UNIVERSIDAD CATÓLICA
de Colombia

IMPLEMENTACIÓN DE UN PROTOTIPO FUNCIONAL DE APRENDIZAJE DE MÁQUINA PARA IDENTIFICAR CORREOS ELECTRÓNICOS DE SPEAR PHISHING

MARIA ALEJANDRA SUAREZ SANCHEZ 67000205
JHINDY HASLEYDE PARDO RODRIGUEZ 67000022

PROGRAMA DE INGENIERÍA DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD CATÓLICA DE COLOMBIA
BOGOTÁ, MAYO
2021

IMPLEMENTACIÓN DE UN PROTOTIPO FUNCIONAL DE APRENDIZAJE DE MAQUINA PARA IDENTIFICAR CORREOS ELECTRÓNICOS DE SPEAR PHISHING

MARIA ALEJANDRA SUAREZ SANCHEZ 67000205
JHINDY HASLEYDE PARDO RODRÍGUEZ 67000022

Trabajo de Grado para optar al título de
INGENIERO DE SISTEMAS Y COMPUTACIÓN

PROGRAMA DE INGENIERÍA DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD CATÓLICA DE COLOMBIA
BOGOTÁ, MAYO
2021

TABLA DE CONTENIDO

INDICE DE FIGURAS

INDICE DE TABLAS

ÍNDICE DE ECUACIONES

AGRADECIMIENTOS

Principalmente agradezco a Dios por brindarme la oportunidad y las fuerzas necesarias para superar todas las dificultades para sacar adelante mi carrera y convertirme en una ingeniera de sistemas y computación, ya que sin él este sueño no hubiera sido posible.
A mis familiares hija, esposo, padres y hermanas por su amor y dedicación, sus consejos, palabras de aliento que me ayudaron a no desfallecer y seguir en esta lucha, creyendo en mí en todo momento y no dudaron de mis habilidades
A mis dos asesores por su tiempo y dedicación para la elaboración de este documento.
A mis docentes y directivos por compartir sus conocimientos, por formarme y prepararme para este sueño, sin perder la esencia de lo que significa ser persona.
A mis compañeros quienes supieron aceptarme para complementarnos con nuestras debilidades y fortalezas e hicieron a un lado nuestras diferencias y me brindaron su amistad, confianza y apoyo, a lo largo de esta carrera y que estuvieron ahí en los buenos y malos momentos.

Muchas gracias a todos, con cariño Jhindy Pardo.

A Dios.

Por darnos la sabiduría y fortaleza de haber culminado esta etapa de estudio con éxito, quiero agradecer a todas aquellas personas que de una u otra forma me apoyaron para alcanzar la meta de mi grado.

A mis principales apoyos y soportes: mis padres, a mis amigos, mi jefe y compañeros de estudios, a mis profesores y guías, y a toda mi familia, quiero hacerles llegar mis palabras de agradecimiento por haberme apoyado hasta hacer realidad mi graduación. ¡Muchas gracias!

Hoy con mi título en mano me invade la felicidad, el orgullo y el agrado por haber cumplido con una de mis más grandes metas: el haberme graduado. De igual forma quiero dar las gracias a todas las personas que me apoyaron en la consecución de mi objetivo. ¡Gracias!

A nuestro director del proyecto de graduación, por su guía, comprensión, paciencia, entrega y valiosos consejos a lo largo del proceso de investigación.

María Alejandra Suarez Sánchez

DEDICATORIA

Con todo el cariño y amor a mis padres y hermanas, por brindarme la oportunidad de estudiar una carrera profesional después de tanto tiempo, a mi esposo e hija por acompañarme en este proceso largo y arduo que hoy este sueño se ha hecho realidad, a mi hija Jessica quien fue mi mayor inspiración y motivación para convertirme en profesional y la cual amo profundamente. A ellos quienes han sido mi mano derecha durante todo este tiempo, les agradezco, por confiar en mí, en creer en mis cualidades y capacidades de que podía culminar con éxito esta etapa académica. A ustedes mi infinito agradecimiento y amor.

Jhindy Pardo

NOTA DE ACEPTACIÓN

_____
_____

_____
_____

Jurado
RAFAEL ACOSTA

FREDY ERNESTO PARDO ANGULO
Asesor

Bogotá, junio del 2021

RESUMEN

Este trabajo tiene como propósito la detección de correos electrónicos Spear Phishing a través de un prototipo web, debido a que las técnicas de ingeniería social son muy usadas hoy en día para robar a los usuarios datos de identidad personal y/o credenciales de sus cuentas financieras, por tal motivo, todas las personas deben implementar una medida para detectar estos ataques de ingeniería social.

Este prototipo realizado, utiliza la técnica de aprendizaje automática como Random forests (árboles de decisión), a través de la recolección de conjunto de datos de correos electrónicos Spear Phishing y Correos electrónicos no Spear Phishing, a la cual se le realizó un proceso de alistamiento, utilizando lematización seguido del uso de Cross-Validation, para finalizar así con un algoritmo en lenguaje de programación Python combinado con el framework flask.

Obteniendo unos resultados satisfactorios en cuanto a la precisión que equivale a un 94% en la detección de correos electrónicos Spear Phishing. Con esto se concluye que la técnica aplicada fue la correcta para dicha detección.

Palabras Claves: Correo electrónico, Seguridad, Inteligencia Artificial, Cibercrimen, identidad

.

ABSTRACT

The purpose of this work is to detect Spear Phishing emails through a web prototype, since social engineering techniques are widely used today to steal personal identity data and / or credentials of their financial accounts. For this reason, everyone people must implement a measure to detect these social engineering attacks.

This prototype, carried out, uses the automatic learning technique such as Random forests (decision trees), through the collection of Dataset of Spear Phishing emails and no-Spear Phishing emails, to which an enlistment process was carried out, using stemming followed using Cross-Validation, thus ending with an algorithm in Python programming language combined with the flask framework.

Obtaining satisfactory results in terms of precision that is equivalent to 94% in the detection of Spear Phishing emails. With this, it is concluded that the applied technique was the correct one for said detection.

Keywords: Email, Safety, Artificial intelligence, Cybercrime, Identity.

1 INTRODUCCIÓN

En la actualidad el crecimiento del número de usuarios de internet ha sido significativo, dando lugar a que personas inescrupulosas saquen ventaja de esta situación, aprovechándose de las vulnerabilidades que tienen los usuarios en el uso de las tecnologías, para la creación de varios métodos o técnicas de estafa.

Uno de ellos es el conocido como Phishing, que es uno de los ciberataques más utilizados. Consiste en la suplantación de identidad a nivel personal o de empresa. Dado que existen muchos tipos de phishing, este proyecto se centrará en el método de spear phishing, que consiste en *"Ataques que se basan en el análisis de los piratas informáticos". Investigan a la víctima, aprenden sobre sus gustos, sus operaciones diarias, información que puedan recopilar, etc. De esta forma logran captar aún más la atención de la víctima."*[1]

Como se mencionaba, este proyecto hablará sobre uno de los métodos más usados para realizar estafas electrónicas, el spear phishing, que es la causa del 57% de los ataques provenientes del sector de la banca.[2]

En Colombia en el 2019 se reportó un 80% de correos fraudulentos personales, teniendo pérdidas que oscilaron entre 300 millones y 5.000 millones de pesos[3], debido a que los usuarios no conocen o no prestan atención a las señales que traen los correos electrónicos bajo la modalidad de spear phishing, logrando que una tercera persona se apropie de sus finanzas.

Este proyecto tiene un enfoque investigativo y social, debido a que se pretende alertar a los receptores del mensaje que están frente a una posible estafa, donde pueden estar en riesgo las finanzas involucradas, al punto de una pérdida total o parcial. Esto a través de una de las ramas de la inteligencia artificial como Machine Learning, el cual permite detectar patrones que se encuentren en correos electrónicos, y así mitigar el impacto que tienen estas estafas en la sociedad, dejando pérdidas millonarias.

---

[1] Alberto Gallego Yuste.2012. DELITOS INFORMÁTICOS: MALWARE, FRAUDES Y ESTAFAS A TRAVÉS DE LA RED Y CÓMO PREVENIRLOS Disponible en:https://e-archivo.uc3m.es/bitstream/handle/10016/16868/pfc_alberto_gallego_yuste.pdf?sequence=1

[2] Hard2bit Cybersecurity. 2019.Disponible en https://hard2bit.com/blog/8-tipos-de-ataque-phishing-que-ponen-en-riesgo-tu-seguridad/.

[3] Sabrina Pagnotta2017. Las víctimas de ciberataques perdieron 1,33 mil millones de dólares en 2016Disponible en https://www.welivesecurity.com/la-es/2017/06/28/victimas-ciberataques-millones-dolares/

2 JUSTIFICACIÓN

Actualmente el avance tecnológico ha conseguido introducirse paulatinamente dentro de la sociedad formando parte indispensable de ella.[4] Sin embargo, se deben soportar riesgos y desafíos con un gran margen de error, como lo son la manifestación de correos fraudulentos que pretenden sustraer la información confidencial del usuario, convirtiéndose en un proceso ilegal fuera de la jurisdicción de los propios sistemas de seguridad y gestión. La técnica más utilizada para conseguir este objetivo, y que se ha convertido en un gran riesgo para la sociedad en general, se denomina Phishing, dado que la simulación perfecta y casi imperceptible de fuentes no fiables conducen al usuario a entrar y otorgar de manera autónoma sus datos, generalmente con fines financieros o bancarios.[5]

Dicho esto, se hace obligatorio determinar los posibles patrones conductuales que utilizan los ciberdelincuentes, mediante la conceptualización de la trascendencia global de éste tipo de delitos y el incremento notable del mismo, con el fin último de desarrollar posibles estrategias de contingencia y políticas organizacionales, desde los sistemas que mitiguen los delitos cibernéticos, y así ayudar a que las personas naturales y/o jurídicas, en los cuales su comunicación masiva se da mediante la viralización de correos electrónicos que no contengan malwares o softwares maliciosos.[6]

La producción de este proyecto se fija en el estudio de los patrones y datos estadísticos que ayuden a la identificación de seguridad de los correos electrónicos, las cifras porcentuales de esta modalidad delictiva, y el desarrollo de herramientas y protocolos necesarios para su prevención en las futuras generaciones. Se pretende que a través de algoritmos de seguimiento basados en técnicas de machine learning se puedan identificar, rastrear y clasificar diferentes elementos y patrones presentes en un correo electrónico enviado por un phishing, para así ofrecer una solución que permita no caer las trampas de los llamados Phishing de clonado/redireccionamiento o Malware-Based Phishing[7]

---

[4]Susana Galeano.2020. Marketing Economía.Disponible en https://marketing4ecommerce.net/usuarios-internet-mundo/

[5] Hard2bit CyberSecurity. 2019.Disponible en https://hard2bit.com/blog/8-tipos-de-ataque-phishing-que-ponen-en-riesgo-tu-seguridad/

[6] Disponible en https://www.welivesecurity.com/la-es/2017/06/28/victimas-ciberataques-millones-dolares/

[7] Sabrina Pagnotta 2017, Welivesecurity .Disponible en de https://www.ibm.com/ar-es/analytics/machine-learning?p1 =Search&p4=43700052827921639&p5 =b&cm_mmc = Search_Google-_-1S_1S-_-LA_ISA-_-%2Bmachine %20%2 Blearning_b&cm_mmca7=71700000065289299&cm_mmca8=kwd-26527633773&cm_mmca9=Cj0 KCQjwreT8BRDTARIsAJLI0KLW1F9REp4nAiURJsUF_yMLIBI18gaPiL7VbvG36CaZEvIH1Z9DScQ aAqKxEALw_wcB&cm_mmca10=429770185321&cm_mmca11=b&gclid=Cj0KCQjwreT8BRDTARIs AJLI0KLW1F9REp4nAiURJsUF_yMLIBI18gaPiL7VbvG36CaZEvIH1Z9DScQaAqKxEALw_wcB&gcl src=aw.ds

3 PLANTEAMIENTO DEL PROBLEMA

De acuerdo al Departamento Administrativo Nacional de Estadística DANE, en Colombia hay 50 millones de habitantes, de los cuales 35 millones de ellos están conectados a internet, es decir el 69%. Si se compara esta cifra con la arrojada a inicios del 2019, se visualiza un incremento del 3.3%. Y de acuerdo con lo afirmado por la página Branch, estos usuarios "*administran su tiempo diariamente de la siguiente manera 9 horas y 10 minutos usando el Internet, 3 horas y 35 minutos usando las redes sociales, 3 horas y 30 minutos viendo televisión, 1 hora y 24 minutos escuchando música a través de servicios de streaming (como Spotify), 0 horas y 53 minutos usando consolas de videojuegos." Como lo dice K. Rosgaby Medina* [8]

Debido al aumento exponencial del uso del internet en el año 2020 tras el brote del Covid-19, las personas incrementaron el uso de esta herramienta en índoles laborales, académicos y sociales en casa. Como consecuencia de ello, y según los puntos de intercambio de Internet (IXP), se ha experimentado hasta un 60% más de tráfico de Internet en comparación con el tráfico previo al brote. Por otra parte, el correo electrónico sigue siendo la herramienta más antigua y útil para intercambiar mensajes de información de datos, imágenes y archivos. Tanta es su importancia, que el diario del portafolio analizó sobre el decreto 806 del 2020, el cual estableció el correo electrónico como el medio de comunicación y notificación en la administración de justicia. Ello ha hecho que este medio sea uno de los más atractivos para realizar ciberataques por parte de delincuentes. Por esta razón la seguridad de la información de los usuarios se ve afectada, debido a que los ataques de phishing han marcado tendencia. Según el informe de investigaciones de violación de datos de Verizon, el 96% de los ataques de phishing llegan por correos electrónicos el 3% se realizaron a través de sitios web y solo el 1% por teléfonos.

Debido a este incremento, la seguridad de la información de los usuarios se ve más comprometida ya que, como lo dijo Baadel y Joan Lu[9] , los usuarios son vulnerables e ingenuos a estos ataques que sin reparo afectan la integridad del uso de herramientas tecnológicas. Debido esta problemática, muchas empresas han optado por mejorar sus esquemas de seguridad, teniendo en cuenta que un factor importante de riesgo es el humano ya que por su imprudencia pueda acceder a Urls o sitios que no sean seguros a través de correos. Por este motivo se pretende detectar los correos fraudulentos enviados por ciberdelincuentes. Ahora bien, una de las tácticas más populares que utilizan los delincuentes en internet se llama vulnerabilidad emocional del ser humano, de modo que permite

[8] K. Rosgaby Medina.Branch.2020. Disponible en: https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-colombia-en-el-2019-y-2020/#:~:text=En%20un%20a%C3%B1o%20(del%202019,crearon%203.4%20millones%20nuevos%20perfiles

[9] Said Baadel and Joan Lu.2019. Data Analytics: Intelligent Anti-Phishing Techniques Based on Machine Learning Disponible en:https://www.worldscientific.com/doi/abs/10.1142/S0219649219500059

el acceso a datos logrando su objetivo. Según Martínez, *"Colombia es uno de los tres mercados latinoamericanos más atractivos para los cibercriminales. Al día, se presentan unos 28.835 ataques de phishing en el país".*[10] A manera de ejemplo de la gran amenaza que significa el phishing, para el mes de abril del 2020 la compañía industrial e informática Google detectó cerca de 18 millones de intentos de ciberataque bajo la excusa de la emergencia sanitaria mundial denominada covid-19 o coronavirus, así mismo, detectó cerca de 240 millones de mensajes considerados como spam o mensajes no solicitados. Mientras que la compañía Microsoft cada día detecta cerca de 60.000 mensajes contenedores de archivos maliciosos, por lo cual, ha advertido acerca de la vulnerabilidad presentada en la emergencia sanitaria en posibles actos delictivos informáticos.[11] Según los estudios de informes de tendencias CDR, finalizando el mes de marzo y a principios del mes de abril los ataques de phishing por correos aumentaron un 37% con falsa información de covid-19, ya que estos cibercriminales usan estrategias de tendencias para llamar la atención de los usuarios, aprovechando así su vulnerabilidad, suplantando otras entidades reconocidas como la DIAN con un 57% la Fiscalía 12%, los organismos de tránsito con 10% policía nacional 9% y ministerio de salud con un 7%.[12]

Los ataques de phishing han crecido durante los últimos años, debido a que los atacantes no dejan de mejorar sus tácticas y reutilizan las más exitosas. En un ataque de *spear phishing* los ciberdelincuentes envían algún tipo de mensaje electrónico; según estudios de State of Phish, el principal impacto que ocasiona es de orden económico, ya que los cibercriminales extorsionan a sus víctimas buscando dinero. Para el 2020 las pérdidas financieras están en un 37%, infección de malware 37%, infección de ransomware 50%, información de cuenta 50% y pérdida de datos con un 55%.[13]

Hoy en día existen herramientas basadas en el uso de Machine Learning como lo es Amazon Fraud Detection que es una herramienta útil de inteligencia artificial. Actualmente se está utilizando el método de aprendizaje automático para detectar con éxito los emails maliciosos y sospechosos "Estos nuevos modelos combinan una amplia variedad de técnicas como, por ejemplo, los análisis de reputación y similitud de las URL, permitiendo generar nuevas advertencias a la hora de hacer click en URL relacionadas con enlaces de phishing y malware. A medida que se

---

[10] Martinez.El tiempo.2019. El cibercrimen no descansa, estas son las proyecciones para el 2020. Disponible en:
https://www.eltiempo.com/tecnosfera/dispositivos/cifras-de-ciberataques-de-2019-y-tendencias-para-el-2020-435508

[11] Marta Juste.2020. Ciberataques: la amenaza aumenta Disponible en:https://www.expansion.com/economía -digital/innovación/2020/05/27/5ecbaee5468aeb0f238b4599.html

[12]Copyright Segunda Edición. 2020.Informe tendencias del cibercrimen en Colombia.Disponible en: https://www.ccit.org.co/wp-content/uploads/informe-tendencias-abril-2020-final.pdf

[13] State of the Phish.An in-depth look at user awareness, vulnerability and resilience. 2020.Disponible en:
https://www.proofpoint.com/sites/default/files/gtd-pfpt-uk-tr-state-of-the-phish-2020-a4_final.pdf

encuentran nuevos patrones, los modelos se adaptan más rápidamente que los sistemas manuales, mejorándolo con su uso a lo largo del tiempo.” [14]

Según el centro de recursos User IT, el 95% de los usuarios no pueden identificar los mensajes de phishing y tan solo un 5% de los usuarios reconocen los correos electrónicos fraudulentos. Debido a estos datos es importante seguir implementando prácticas para reducir los ataques de en correos electrónicos.

[14] CSO Computerworld Gmail actualiza su sistema de seguridad Disponible en: https://cso.computerworld.es/tendencias/gmail-actualiza-su-sistema-de-seguridad

3.1 PREGUNTA DE INVESTIGACIÓN

¿De qué manera se puede automatizar el proceso de detección de phishing en correos electrónicos?

3.2 ALCANCES Y LIMITACIONES

El prototipo propuesto se basa sólo en la detección de correos electrónicos Spear Phishing utilizando la técnica de inteligencia de artificial Arboles de decisión.

El tamaño del Dataset depende de muchas variables las cuales están atadas a la actualidad de los datos y el riesgo que conlleva a conseguir ya que en las entidades bancarias o empresariales no existe un repositorio de dicha información o si lo posee no es fácil acceder a ello.

El prototipo ejecutado sólo realiza detecciones de correos Spear Phishing en idioma anglosajón (inglés).

La velocidad con la que se ejecuta el prototipo realizado está ligada a las características del hardware, dado que depende mucho del procesamiento de éste para ejecutar el prototipo y detectar así si es un correo Phishing o no. Esto podría involucrar demoras en hacer la detección.

El prototipo web sólo está diseñado para ser ejecutado en ordenadores, mas no está habilitado para sitio web.

4 OBJETIVOS

4.1 OBJETIVO GENERAL

Implementar un prototipo funcional para la detección de Spear phishing en correos electrónicos mediante técnicas de aprendizaje automático.

4.2 OBJETIVOS ESPECÍFICOS

Construir un conjunto de datos para el entrenamiento de un algoritmo basado en aprendizaje de máquina que realice la detección de correos maliciosos Spear phishing.

Diseñar una arquitectura de software del prototipo funcional que permita identificar los correos maliciosos con características Spear phishing.

Desarrollar pruebas de concepto web del prototipo funcional para medir el desempeño del algoritmo de clasificación de Spear phishing en correos electrónicos.

5 MARCOS DE REFERENCIA

5.1 MARCO CONCEPTUAL

En el marco conceptual se detallan los modelos de teoría, los conceptos, argumentos e ideas que se desarrollaron en el trabajo de grado.

5.1.1 **Correo Electrónico**

Sistema que permite el intercambio (enviar y recibir) mensajes entre distintas computadoras interconectadas a través de una red.[15]
Estos mensajes consisten en la transferencia de (texto, imágenes, videos, documentos, reuniones, grabación de audios y link). Fue diseñado para la comunicación entre dos personas o incluso el mismo mensaje se puede enviar a múltiples destinatarios como se desee sin importar que todos los remitentes sea de diferentes lugares tanto del país como de ciudad.
La estructura del mensaje electrónico se basa en un destinatario (se pueden adicionar varios correos de usuarios), el remitente, el asunto (título que se le da al mensaje), y el cuerpo del texto. Este último puede incluir archivo de firma con los datos del remitente.

Figura 1. Estructura Correo Electrónico

Tomado de: https://sites.google.com/site/ticsmielca/estructura-de-un-mensaje-de-correo-electronico

Además, los correos electrónicos tienen una ventaja y es que si el mensaje no llega al destinatario (porque no exista esa dirección o esté mal escrita) este devuelve una alerta al remitente con un mensaje aclaratorio.[16]

[15] Keyla, Calameo, Disponible en https://es.calameo.com/read/004971302896b8f50921a
[16] Proyecto hola orientación, Correo electrónico Tomado de: https://www.uv.mx/personal/rcordoba/files/2014/11/Correo_electronico.pdf

5.1.2 **UBE (Unsolicited Bulk Emails)**

Nombre formal que se le da a los correos electrónicos no deseados que son enviados a usuarios que no los han solicitado.[17]
La mayoría de estos correos son anuncios publicitarios y constituyen una amenaza para los usuarios debido a que puede ingresar un virus o ataques phishing.

5.1.3 **Adjuntos de correo electrónico**

Los archivos que llegan con el correo electrónico forman parte del mismo paquete, el correo puede ir sin un archivo adjunto, pero jamás al revés. Los correos institucionales tienen un límite de capacidad para adjuntar archivos, pero en los correos gratuitos se encuentran limitados según el proveedor servicios.[18]

5.1.4 **Ventanas emergentes**

La principal característica de las ventanas emergentes o desplegables es que llaman la atención y dirigen al usuario, aunque sea durante unos segundos, a un lugar concreto dentro del sitio que está navegando una persona. Son el resultado de una acción llamada pop-up que se activa cuando el usuario elige una opción o algún componente dentro de la página web en revisión. Son uno de los métodos intrusivos que se utilizan para vulnerar la seguridad de una persona en Internet.[19]

Ahora en los siguientes conceptos se investigará acerca de cómo la población realiza técnicas de engaño a usuarios para obtener información.

5.1.5 **Ingeniería social**

Método utilizado por los atacantes para engañar a los usuarios informáticos, los cuales realizan una acción la cual produce consecuencias negativas, como la divulgación de información personal. Los ataques de phishing son los que con mayor frecuencia usan estas tácticas con el fin de atacar las vulnerabilidades humanas para conseguir un beneficio como las claves y datos confidenciales del usuario atacado, esto con el fin de usar dicha información para acceder de manera abusiva a un sistema informático.[20]

5.1.6 **Suplantación**

---

[17] Copyright ©.2020.What is an Unsolicited Bulk Email? – Basic Disponible en: https://sendpulse.com/support/glossary/unsolicited-bulk-email
[18] 1&1 IONOS ESPAÑA S.L.U.2020. ¿Qué son archivos adjuntos de correo? Disponible en https://www.ionos.es/ayuda/correo/glosario-explicaciones-sobre-conceptos-y-temas-importantes/archivos-adjuntos-de-correo/
[19] Software DELSOL.Popups o ventanas emergentes. Disponible en https://www.sdelsol.com/glosario/popups-o-ventanas-emergentes/
[20] Sergio Arcos Sebastián, Ingeniería social Psicología aplicada a la seguridad informática tomado de: https://idoc.pub/documents/ingenieria-social-1430g9713j4j

Es una de las técnicas más comunes e importantes cuando se quiere realizar ingeniería social, pues esta consta de realizar un estudio detallado con el fin de hacerse pasar por una persona o red, buscando engañar al usuario para que revele información personal como lo es contraseña, número de teléfono, tarjetas de crédito y tarjeta débito de la víctima. Esta técnica intenta engañar al usuario para que realice una acción que sólo haría con una entidad confiable. El fin es que el sitio reemplazado simule el funcionamiento de esta, un ejemplo es la página de bancos o entidades gubernamentales.

5.1.7 **Ciberataque**

Es cualquier intento de obtener acceso no autorizados a una o más computadoras o redes. El objetivo de los atacantes es desactivar, interrumpir, destruir o controlar los sistemas informáticos, así como también bloquear, eliminar, manipular o robar los datos almacenados en esto sistemas. Los ataques se dividen en dos grupos los dirigidos y los no dirigidos.[21]

5.1.7.1 Los ataques dirigidos

Este es conocido porque el atacante tiene un interés especifico, ya sea un negocio o que le hayan pagado para realizar el ataque. Este ataque dirigido tiene que llevar un estudio de meses o incluso años para encontrar la manera de atacar al usuario. Este ataque se considera el más dañino, porque es diseñado especialmente para dañar el sistema, proceso, entidad o persona los ataques que están en el grupo de dirigidos son:

Spear-phishing: Envío de correo electrónico a una persona o entidad en específico que contiene un archivo adjunto con software malicioso o enlace de descarga.
Implementación de una botnet: para realizar un ataque DDOS (denegación de servicio distribuida).
Subvertir la cadena de suministro: atacar equipos o software que se entregan a la organización.

5.1.7.2 Ataques no dirigidos

Estos ataques son enviados a varios dispositivos, servicios o usuarios. En este tipo de ataque no importa quién es la víctima, ya que saben que existen dispositivos o máquinas vulnerables. Es por esto que utilizan técnicas para aprovechar la apertura de internet. Entre estos tenemos:

Phishing: Envío de correo electrónico a gran número de personal solicitando información confidencial (datos bancarios) o adjuntado un enlace para que el usuario acceda a un sitio web falso.

[21] (15 de Octubre del 2015).National Cyber security Centre, tomado de: https://www.ncsc.gov.uk/information/how-cyber-attacks-work

Water holing: Configurar un sitio web confiable para explorar a los usuarios visitantes
Ransomware: Tipo de programa dañino que restringe el acceso a determinadas partes o archivos del sistema operativo. El fin de este ataque es solicitar rescate a cambio de quitar la restricción realizada por el atacante.

5.1.8 **Hackers**

Se define como la persona experta en algunas ramas de la informática que usa este conocimiento con fines ilegales. Esta persona se dedica a detectar fallas de seguridad en los sistemas informáticos a través de diferentes técnicas, están pueden ser malas o buenas. Entre las malas están instalar un programa maligno, robar o destruir datos, interrumpir un servicio y más. y entre las buenas están razones éticas como encontrar vulnerabilidades de software para poder ser resueltas, con el fin de solucionar dichas debilidades de seguridad antes de ser explotadas. Existen muchos tipos de hackers, pero nos enfocaremos en tres tipos:[22]

5.1.8.1 Black Hat Hackers

Conocidos como los hackers de sombrero negro, estos son conocidos como personas inescrupulosas que rompen la seguridad de una computadora y crean virus, interrumpiendo los sistemas de seguridad de computadoras, colapsando servidores, entrando a zonas restringidas e infectando redes apoderándose de ellas.

5.1.8.2 White Hat Hackers

Conocidos como los hackers de sombrero blanco, estos son los hackers éticos que tienen como función para encontrar vulnerabilidades. Se centran en asegurar y proteger los sistemas de tecnologías de información y comunicación. Algunos son consultores de seguridad, trabajan para alguna compañía en el área de seguridad informática protegiendo los sistemas.

5.1.8.3 Gray Hat Hackers:

También llamados Hackers de Sombrero Gris, estos hackers juegan a ser los buenos y los malos. Por lo general no hackean para beneficio personal ni tienen intenciones maliciosas, pero pueden estar dispuestos a comprometerse técnicamente crímenes durante el curso de sus hazañas tecnológicas con el fin de lograr una mayor seguridad.

[22] Carlos Alberto Flores Quispe, Universidad Mayor de San Andrés Carrera de Informática Análisis y Diseño de Sistemas de información, TIPOS DE HACKERS

5.1.9 **URL**

Son las siglas en inglés de Uniform Resource Locator, que en español significa Localizador Uniforme de Recursos. Como tal, el URL es la dirección específica que se asigna a cada uno de los recursos disponibles en la red con la finalidad de que estos puedan ser localizados o identificados.[23]
Sirve para encontrar aquello que se está buscando en una página, un sitio, un archivo, documento entre otros, e. Estas direcciones se componen de “https” que sería el protocolo de acceso para las páginas de internet, “www” dirección de recurso, “.com” es un tipo de dominio.

5.1.10 **Fraude**

El fraude es una de las maneras más antiguas de corrupción, tiene como herramienta el engaño a base de mentiras. En informática es una herramienta muy importante que usan los ingenieros sociales para obtener información de una persona para manipular datos u obtenerlos de forma ilegítima, para poder lucrarse de ellos.

Algunas clasificaciones de fraude que pueden afectar a unas personas son:

Fiscal o tributaria, que manifiesta la evasión del pago de los impuestos como lo indica el Artículo 305 del código penal.[24]
Electoral, es el engaño a la ciudadanía sobre los resultados obtenidos durante unas elecciones.
Laboral o Empresarial, es cuando una persona aprovecha la ocupación o empleo para el enriquecimiento personal, a través del mal uso de los recursos o activos de una compañía.
Fluidos, es cuando una persona de manera clandestina o que altera el sistema de control, se apropia de energía eléctrica, agua, gas o señal de telecomunicaciones, como lo indica el Articulo 256 del código penal[25]
Bancario, es cuando un funcionario de alguna entidad realiza practicas ilegal para obtener información y datos privados de terceros para suplantaciones o hurto de dinero de la persona afectada[26]
Informático o virtual, es una actividad ilícita que tiene como objetivo causar daño, pérdida de información a través de notificaciones, actividades en redes sociales.
Electrónico o telefónico, es la comunicación de una persona a través de una llamada donde la manipulación de algún tipo de información obtiene un beneficio económico.

---

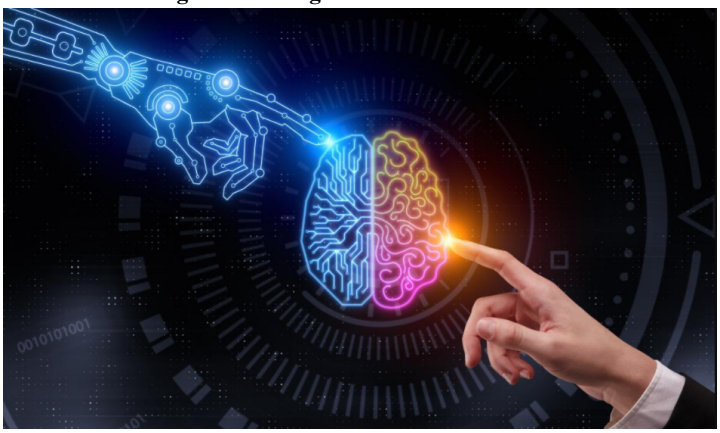
[23] Tecnología e Innovación, Disponible en https://www.significados.com/url/#:~:text=URL%20son%20las%20siglas%20en,significa%20Localizador%20Uniforme%20de%20Recursos.&text=As%C3%AD%2C%20hay%20un%20URL%20para,por%20primera%20vez%20en%201991
[24] Tomado de http://www.secretariasenado.gov.co/index.php/constitucion-politica
[25] Tomado de http://www.secretariasenado.gov.co/index.php/constitucion-politica
[26] Tomado de https://www.larepublica.co/fraude-bancario

En seguida se tratarán los conceptos acerca de los temas relacionados con inteligencia artificial que son importantes para el desarrollo de este trabajo.

5.1.11 **Inteligencia artificial**

El campo científico de la informática que se centra en la creación de programas y mecanismos que pueden mostrar comportamientos considerados inteligentes.[27]

Figura 2. Inteligencia artificial



Tomado de: “Inteligencia artificial-Computer Hoy”, 24 de Agostos del 2019, https://computerhoy.com/reportajes/tecnologia/inteligencia-artificial-469917

Es la combinación de algoritmos con el propósito de crear máquinas que realicen ciertas actividades que realiza el ser humano

Algunos expertos en la ciencia de la computación indican que existen diferentes tipos de la inteligencia artificial[28]

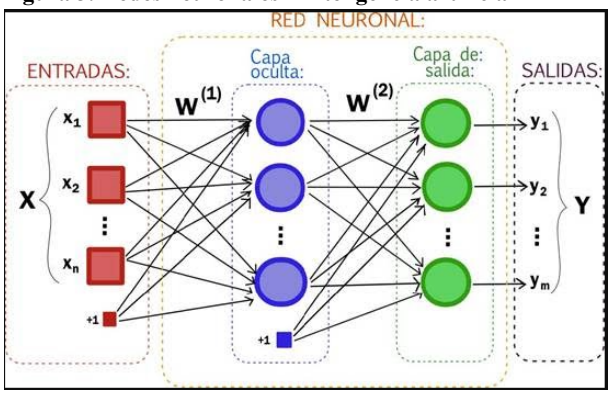
5.1.11.1 Redes neuronales

Son sistemas que piensan como el ser humano, como por ejemplo en actividades de toma de decisión, resolución de problemas y aprendizaje.

[27] SALESFORCE LATINOAMÉRICA.2017. Inteligencia Artificial ¿Qué es? Disponible en: https://www.salesforce.com/mx/blog/2017/6/Que-es-la-inteligencia-artificial.html#:~:text=La%20Inteligencia%20artificial%20es%20el,m%C3%A1quinas%20piensan%20como%20seres%20humanos%E2%80%9D.

[28] Tomado de https://www.iberdrola.com/innovacion/que-es-inteligencia-artificial

Figura 3. Redes neuronales – inteligencia artificial

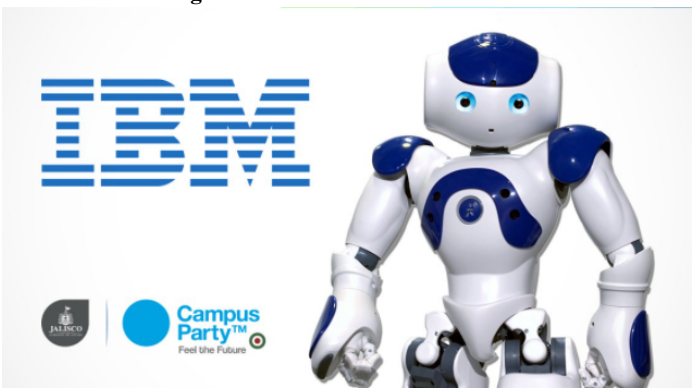


Tomado de: Inteligencia artificial-Unidad 4 Redes Neuronales, https://sites.google.com/site/mayinteligenciartificial/unidad-4-redes-neuronales

5.1.11.2 Robots

Es una computadora que realiza tareas similares a las de un ser humano. Un ejemplo es i-Watson, creado por científicos de IBM, en el cual desarrollaron un sistema de inteligencia artificial que pretende ser capaz de emular y superar al cerebro humano en tareas como jugar ajedrez, organizar documentos entre otras.[29]

Figura 4. Robot i-Watson



Tomado de: Digital TOO, 2019, https://www.digitaltoo.com/2016/03/17/te-presentamos-connie-la-conserje-robot-hilton-e-ibm

5.1.11.3 Pensamiento lógico racional

Es la encargada de lograr que una máquina pueda percibir, razonar y actuar. También se llaman sistemas expertos o agentes inteligentes.

[29] 8 de junio del 2016 / el diario el pais/ autor JAVIER SAMPEDRO / https://elpais.com/elpais/2016/06/08/opinion/1465383749_599768.html

5.2 MARCO TEÓRICO

En el marco teórico se describen los métodos para realizar el proceso de detección de spear phishing en correos electrónicos.

5.2.1 **Phishing**

Es una estafa en línea en la que los cibercriminales envían mensajes fraudulentos que parecen provenir de una fuente legítima. Estos mensajes están diseñados para engañar usuarios los cuales pueden llegar a ingresar información confidencial ejemplo (número de cuentas, PIN, contraseñas).

Los correos electrónicos incluyen enlaces o archivos adjuntos que, una vez el usuario da click, robarán información confidencial o puede llegar infectar los computadores o malware. Los cibercriminales roban está información y algunas veces extorsionan a los usuarios para devolver la información recogida o para robar los datos obtenidos.[30]

Tipos de Phishing: Existen diferentes tipos de phishing hoy en día, de los cuales el principal objetivo es obtener información personal ya sea de empresas y personas, esto con el fin de sobornar y tener beneficios económicos. Los phishing más comunes son:

5.2.1.1 Phishing Spear:

Es el tipo de phishing más usado para robar información confidencial. Este tipo de phishing va dirigido a un tipo de persona, empresa u organizaciones específicas, además de usar tácticas de envío de estos mensajes usando información de entidades conocidas como lo son la DIAN, bancos, fiscalía entre otros. A menudo roban datos para fines maliciosos.

5.2.1.2 Phishing Vishing

Este es un tipo de estafa que es comúnmente realizada por teléfono por medio de IP. Este tipo de phishing utiliza el mismo patrón de engaño, debido a que los cibercriminales crean un mensaje de estafa, como por ejemplo una urgencia, para convencer a la víctima de divulgar datos personales.

Estas llamadas a menudo suelen ser realizadas a través de entidades o identificaciones falsas, que parecen de una fuente confiable, pero son los cibercriminales los encargados de crear escenarios.

[30] la comisión federal de comercio Información para consumidores, Disponible en https://www.consumidor.ftc.gov/articulos/como-reconocer-y-evitar-las-estafas-de-phishing

5.2.1.3 Phishing Ballenero:

El nombre de este tipo de phishing viene dado debido a su alto nivel. Este intento de robar información confidencial suele ser más sofisticado que los demás tipos de phishing enviado por correo y es uno de los tipos de phishing más difíciles de detectar, ya que estos correos van dirigidos a un usuario específico que tenga influencia en la organización y así se lograr el objetivo de recolectar la información y atacar la empresa.

5.2.1.4 Phishing Smishing

Este tipo de phishing usa mensajes de teléfono móvil (SMS). Está técnica comenzó en Europa y Japón en el año 2006 y desde entonces se ha extendido en el mundo. Usa las mismas tácticas de engaño, pero es por medio de SMS, tratando engañar a los usuarios para obtener sus datos personales, número de cuentas bancarias y contraseñas.

5.2.1.5 Clonar Phishing

Como su nombre lo indica su táctica consiste en clonar un correo legítimo y enviarlo. El mensaje clonado está dirigido con contenido idéntico pero malicioso, y parece provenir de un remitente original.[31]

5.2.2 **Cibercrimen**

Se conoce como una actividad delictiva que afecta a una computadora o dispositivo de red.

Estos cibercrímenes son organizados por una persona o grupo de personas cibercriminales, también conocidos como hackers, que pretenden extorsionar a usuarios. Estos cibercriminales usan tácticas de tendencia para llamar la atención de usuarios. Un ejemplo es la suplantación de entidades (bancos, entidades bancarias) reconocidas, realizando correos electrónicos que parecen y se ven idénticos a los mensajes que provienen de la empresa legitima y que contienen una llamada de acción urgente con solicitud para hacer click en el enlace. Una de las tácticas más recientes y favoritas de los cibercriminales es la de enviar un correo electrónico a un cliente que pretende ser de la compañía de la tarjeta de crédito del usuario. El correo electrónico dice típicamente que un cargo por una compra de pornografía infantil está a punto de aparecer en el extracto de la tarjeta de crédito del usuario y le solicita que haga click en un enlace si desea cancelar el cargo.[32]

---

[31]Ivan Belcic.2020. Guía esencial del phishing: cómo funciona y cómo defenderse. Disponible en: https://www.avast.com/es-es/c-phishing

[32] Robert Philip ZagerWilliam AmesJosé Jesús Picazo, Jr. Nageshwara Rao VempatyVikram DuvvooriChris David Trytten. 2004 Techniques for to defeat phishing

El cibercrimen se divide en dos categorías. La primera es la actividad delictiva que va dirigida a la computadora, ya que estos cibercriminales pueden infectarla con virus o malware para dañar o robar datos y documentos en las computadoras. La otra categoría es la actividad delictiva que utilizan computadoras para cometer otros delitos.

Existen diferentes tipos de cibercrimen algunos de estos son:

- Fraude por correo electrónico e Internet.
- Fraude de identidad (en caso de robo y uso de información personal).
- Robo de datos financieros o de la tarjeta de pago.
- Robo y venta de datos corporativos.
- Ciberextorsión (amenazar con un ataque para exigir dinero).
- Ataques de ransomware (un tipo de ciberextorsión).
- Ciberespionaje (en el que los hackers acceden a datos gubernamentales o empresariales).[33]

5.2.3 **Machine Learning:**

Es un algoritmo del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente, mediante un modelo de entrenamiento de datos.

El aprendizaje automático o machine learning en inglés, es el subcampo dentro de las ciencias de la computación especializado en el reconocimiento de patrones complejos en conjuntos de datos. Esta programación se realiza por repeticiones, es decir, que ejecuta una y otra vez la misma operación. La principal característica del aprendizaje automático es que sus programas consiguen extraer de forma autónoma.

Con este tipo de algoritmos interactuamos día a día. Unos de los modelos que utiliza esta técnica es la cámara de fotos del móvil cuando reconoce una cara o cuando se utiliza una aplicación de traducción automática.[34]

Existen tres grupos de algoritmos en aprendizaje automático:

Algoritmos supervisados

[33] 2020. AO Kaspersky Lab. Consejos para protegerse contra el cibercrimen. Disponible en: https://latam.kaspersky.com/resource-center/threats/what-is-cybercrime
[34] Carlos González García, Ciencia Cognitiva Tomado de: http://www.cienciacognitiva.org/files/2017-20.pdf

Este grupo de algoritmos utilizan un conjunto de datos de entrenamiento etiquetados. Estos algoritmos se procesan para realizar predicciones sobre los mismos, corrigiéndolas cuando son incorrectas. El proceso de entrenamiento continúa hasta que el modelo alcanza un nivel deseado de precisión.

Algoritmos semi - supervisados

En este grupo se combinan tanto datos etiquetados como no etiquetados para generar una función deseada o clasificador. Este tipo de modelos deben aprender las estructuras para organizar los datos, así como también realizar predicciones.

Algoritmos no supervisados

Con el algoritmo no supervisado, el conjunto de datos no se encuentra etiquetado y no se tiene un resultado conocido. Por ello deben deducir las estructuras presentes en los datos de entrada. Lo puede conseguir a través de un proceso matemático para reducir la redundancia sistemáticamente u organizando los datos por similitud.[35]

5.2.4 **Pandas**

Paquete de Python que proporciona estructuras de datos rápidas, flexibles y expresivas diseñadas para hacer que el trabajo con datos "relacionales" o "etiquetados" sea fácil e intuitivo.[36] Panda lleva tareas importantes, como por ejemplo el alineado de datos para su comparación. Es la más usada para ciencia de datos y aprendizaje automático, ya que ofrece una estructura de datos expresiva y flexible que nos facilita la manipulación y el análisis de estos. Si se utiliza esta librería en el desarrollo se podrá lograr modelar, cargar, analizar, manipular y preparar los datos.

Panda dispone de 3 estructura de datos básicas, las cuales son: la estructura de serie, estructura de cubo y la estructura DataFrame que es una colección ordenada de columnas con nombres y tipos, estos son parecidos a una tabla de base de datos (Excel), donde una sola fila representa un único caso (ejemplo), y las columnas representan atributos particulares[37]. Se puede acceder a sus elementos mediante los nombres de las filas y las columnas. Está también la estructura en serie que es de una dimensión. Sus elementos tienen que ser del mismo tipo. Por último la de panel, que equivale a tres dimensiones de datos.[38]

---

[35] Tratamiento masivo e Datos Utilizando técnicas e Machine learning Tomado de: https://digital.cic.gba.gob.ar/bitstream/handle/11746/5603/11746_5603.pdf-PDFA.pdf?sequence=1&isAllowed=y

[36] Aprendizaje Pandas; Free unaffiliated eBook created from Stack Overflow contributors. Tomado de: https://riptutorial.com/Download/pandas-es.pdf

[37]Abder-Rahman Ali, Introducción a pandas Tomado de: https://riptutorial.com/Download/pandas-es.pdf

[38] 04 octubre 2020. La librería Pandas Tomado de: https://aprendeconalf.es/docencia/python/manual/pandas/

Figura 5. Importar librería Pandas en Python

```
import pandas as pd
```

5.2.5 **Sklearn**

Scikit-learn es una librería de Python para aprendizaje automático y análisis de datos. Es fácil de usar y contiene gran cantidad de técnicas de aprendizaje automático para implementar. Con esta librería se puede realizar aprendizaje supervisado y no supervisado[39]. Es útil para resolver problemas ya sea de clasificación (identificar a qué categoría pertenece un objeto) de regresión (predecir un atributo de valor continúo asociado con un objeto) de agrupación (agrupación automática de objetos similares en conjuntos).[40]

5.2.6 **Tf-idf**

Las siglas en inglés de Term frequency – Inverse document frequency, (Frecuencia de término – frecuencia inversa de documento). Analiza la importancia de ciertas palabras de un texto para luego hacer un cálculo inverso y descartar aquellas excesivamente repetidas. Su mayor aplicación son los sistemas de recuperación de información y minería de texto.

TF - Frecuencia de términos: es un valor que representa la cantidad de veces que una palabra que se repite dentro de un documento.

Ecuación 1. Fórmula TF

$$TF = \frac{t \in d}{T \in d}$$

Cantidad de veces que una palabra aparece en un documento

Cantidad de palabra totales del documento

IDF - Frecuencia inversa de documento: Analiza la relevancia de una palabra en base al número de veces que se repite en un documento.

Ecuación 2. Fórmula IDF

$$IDF = x\,log(\frac{D}{\{d \in D : t \in d\}})$$

Todos los documentos

La suma de todas las veces que aparece la palabra en todos los documentos

Ahora, para calcular el TF IDF, se aplican las fórmulas anteriormente expresadas, lo cual arrojará el peso de una palabra en un documento.[41]

Ecuación 3. Fórmula TFIDF

$$TF - IDF = TF * IDF$$

---

[39] 10 octubre 2020.Jose Martínez Heras. IArtificial.net 15 Librerías de Python para Machine Learning Tomado de: https://www.iartificial.net/librerias-de-python-para-machine-learning/

[40]scikit-learn Machine Learning in Python. Tomado de: https://scikit-learn.org/stable/

[41] Tomado de Publisuites, Eric Seguí Parejo, https://www.publisuites.com/blog/tf-idf/#Que_es_el_TF_IDF

5.2.7 **Fórmula F1 Score**

Esta fórmula consigue optimizar los clasificadores, aun sin estar balanceado el conjunto de datos, debido a que no se afectará. Esto ayuda a mantener los falsos positivos y los falsos negativos, gracias a una buena métrica entre Precisión y Recall.[42]

Ecuación 4. Fórmula de F1-Score

$$F1 = 2\frac{precision*recall}{precision+recall}$$

Recall o Exhaustividad: La cantidad de casos clasificados como verdaderos positivos sobre todo lo que realmente era positivo
Precisión: La cantidad de casos verdaderos positivos sobre la cantidad total de todo lo que dijiste que era positivo

Ecuación 5. Precisión

$$precision = \frac{TP}{TP+FT}$$

Ecuación 6. Fórmula recall

$$recall = \frac{TP}{TP+FN}$$

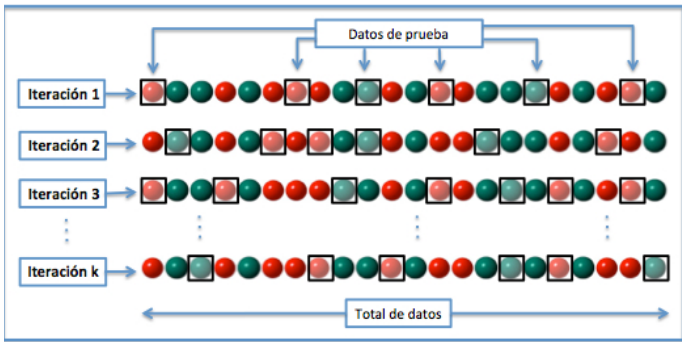
TP- Verdadero positivo: la predicción es un ejemplo positivo, el real es un ejemplo positivo
FP- Falso positivo: la predicción es positiva, la real es negativa
TN- Verdadero negativo: la predicción es negativa, la real es negativa
FN-Falso negativo: la predicción es un ejemplo positivo, el real es un ejemplo negativo

5.2.8 **Cross-Validation o Validación Cruzada:**

Es una técnica que se usa para evaluar los resultados de análisis estadísticos, que consiste en dividir los datos en varios conjuntos, pero cada subconjunto se divide en dos subconjuntos que son los datos de prueba y los datos de entrenamiento. Existen diferentes modelos de validación cruzada.

[42] Explicación alternativa para accuracy, precision, recall y f1-score, {en línea}, Mayo del 2019, Disponible en: https://steemit.com/spanish/@waster/explicacion-alternativa-para-accuracy-precision-recall-y-f1-score

Figura 6. Validación cruzada

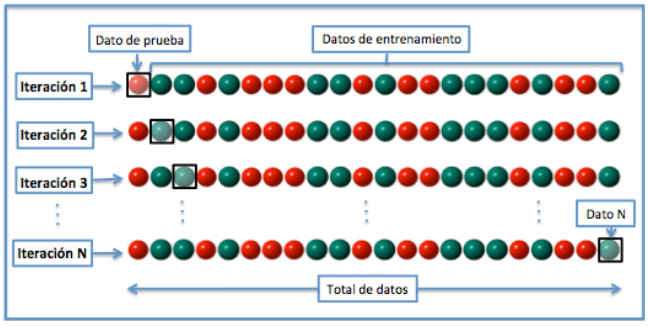


Tomado de https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada

Validación LOOCV:

Este modelo permite reducir la variabilidad que se ocasiona si se divide los datos aleatoriamente. Debido a que el proceso de LOOCV implementará todos los datos disponibles tanto como entrenamiento como prueba, y los resultados de LOOCV son totalmente reproducibles.
Es un método de validación muy desarrollado, ya que se puede ajustar a cualquier tipo de modelo. Sin embargo, una de sus desventajas es que, al emplearse todas las observaciones como entrenamiento, se puede estar incurriendo en un sobreajuste (overfitting).

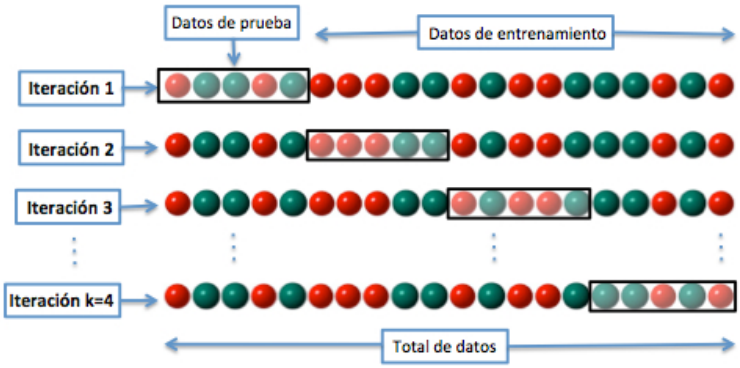
Figura 7. Validación cruzada LOOCV



Tomado de:https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada#/media/Archivo:Leave-one-out.jpg

Validación de k iteraciones o K-Folds: Es uno de los modelos menos sesgados para realizar las comparaciones. Los datos de la muestra se dividen en k iteraciones, y de cada iteración se obtiene un subtipo de datos de pruebas que son diferentes en cada iteración, como se muestra en la figura 8. Este modelo será el utilizado en este proyecto.

Figura 8. Validación Cruzada de K iteraciones

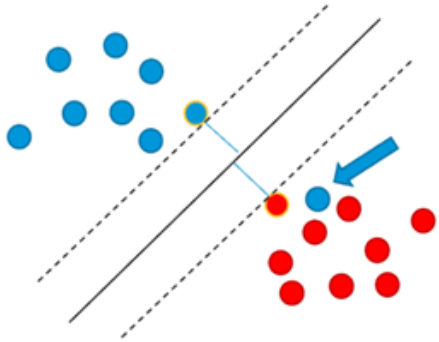


Tomado de: Wikipedia, https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada

5.2.9 **Máquinas de Vectores de Soporte (SVM):**

También son conocidas con Support Vector Machines (SVM), se pueden usar tanto para regresión como para clasificación. Consisten en una implementación del uso de algoritmos matemáticos y estadísticos para poder encontrar patrones de los datos. Se pueden usar en la clasificación binaria, clasificación multiclase, regresión, selección de variables, identificación de casos anómalos y clustering. Sus aplicaciones van desde la detección de rostros, detección de intrusos, clasificación de correos electrónicos hasta artículos de noticias o uso en páginas web.

Figura 9. Máquinas de vectores de soporte (SVM)



Tomado de: IArtificial.net, José Martínez Heras, 28-05-2019,
https://www.iartificial.net/maquinas-de-vectores-de-soporte-svm/

Algunos términos que se deben tener en cuenta son los siguientes para comprender SVM son:

Vector de soporte: Son los puntos de datos más cercanos al hiperplano
Hiperplano: es un plano de decisión que separa entre un conjunto de datos
Margen: espacio entre las dos líneas en los puntos más cercanos de la clase

La SVM construye un hiperplano en un espacio multidimensional con el fin de separar las distintas clases, generando un hiperplano óptimo de forma interactiva que se utiliza para minimizar el error. Su función principal es segregar un conjunto de datos de la mejor manera posible, obteniendo la distancia entre los puntos más cercanos como el margen. El objetivo es seleccionar un hiperplano con el máximo margen posible entre vectores de soporte en el conjunto de datos.

Expresiones matemáticas:

$$w^T x + b = 0 \quad (1)\ \text{Hiperplano solución}$$

W = Vector ortogonal al hiperplano

b = coeficiente de intersección

$$h^+ \rightarrow w^T x_i + b = +1 \quad (2)\ \text{Hiperplano positivo}$$

$$h^- \rightarrow w^T x_i + b = -1 \quad (3)\ \text{Hiperplano negativo}$$

$$\frac{2}{|w|} \quad (4)\ \text{Margen}$$

$$f(x) = (w^T x + b) \quad (5)\ \text{Clasificación SVM}$$

Cuando los datos no se pueden separar linealmente se usa la función llamada Kernel, en la cual se hace un cambio de espacio; las funciones polinómicas y la función de base radial también ayudan a separar los datos (x_i , y_i)

5.2.10 **Random forest**

También conocido como Bosque Aleatorio, es una combinación de árboles de predicción de manera que cada árbol depende de los valores de un vector aleatorio muestreado de forma independiente y con la misma distribución para todos los árboles del bosque.

Random forest da una estimación de qué variables son importantes en la clasificación. Genera también una estimación objetiva de los errores generalizados para los bosques. Converge en un límite a medida que aumenta el número de árboles en el bosque. El error de generalización de un bosque de clasificadores de árboles depende de la fuerza de los árboles individuales en el bosque y la correlación entre ellos.

Random Forest funciona de la siguiente manera:

Se seleccionan k columnas de las m totales (siendo k menor a m) y creamos un árbol de decisión con esas k características.

Se crean n árboles variando siempre la cantidad de k columnas y también se pueden variar la cantidad de muestras que se pasan a esos árboles. Esto es conocido como “bootstrap sample”
Se toma cada uno de los n árboles y se les pide que hagan una misma clasificación. Se guardan los resultados de cada árbol obteniendo n salidas.
Se calculan los resultados obtenidos para cada “clase” seleccionada y se considera a la más escogida como la clasificación final de nuestro “bosque”.

Figura 10. Randon Forest

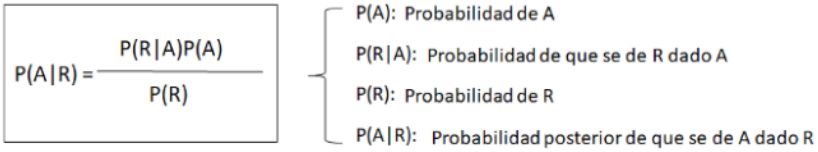
Fuente: Autores

5.2.11 **Naive Bayes:**

Es una clase especial de algoritmos de clasificación de Aprendizaje Automático, o Machine Learning. Se basan en una técnica de clasificación estadística llamada “teorema de Bayes”.

Este modelo bayesiano de probabilidad condicionada se representa como:

Figura 11. Cálculo de modelos Naive Bayes



Tomado de:
https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fudamentos-e-implementaci%C3%B3n-4bcb24b307f

Es decir, la probabilidad de que se dé el caso A dado B es igual a la probabilidad de la intersección de A con B (A∩B) partido la probabilidad de B.

5.2.12 **Framework:**

Se refiere a una estructura de software compuesta de componentes personalizables e intercambiables para el desarrollo de una aplicación. Un framework se puede decir que es una aplicación genérica incompleta y

configurable a la que se le pueden añadir las últimas piezas para construir una aplicación concreta.[43]

Framework Web:

Es un conjunto de clases en java y descriptores y archivos de configuración en XML que componen un diseño reutilizable para así facilitar y agilizar el desarrollo de sistemas Web.

5.2.13 **JavaScript:**

Es un lenguaje de programación de tipo interpretado en el cliente por el navegador al momento de cargar la página. También conocido como lenguaje orientado a objetos, los programas escritos con JavaScript se pueden probar directamente en cualquier navegador sin necesidad de procesos intermedios y es muy utilizado principalmente para crear páginas web dinámicas.[44]

5.2.14 **CSS**

Son las siglas en inglés de Cascading Style Sheets (Hojas de Estilo en Cascada). Consiste en un lenguaje de reglas en el cual se puede configurar el aspecto de las páginas web en el navegador. El navegador contiene los estilos CSS a los elementos seleccionados para exhibirlos correctamente. Es una de las tres principales tecnologías web, junto con HTML y JavaScript.[45]

5.2.15 **HTML**

Lenguaje de marcado de hipertexto o HyperText Markup Language por sus siglas en inglés. Es utilizado para el desarrollo de páginas en internet. HTML sirve para indicar cómo va ordenado el contenido de una página y esto lo logra por medio de las marcas de hipertexto, las cuales son etiquetas conocidas en inglés como tag (<>). Estas son útiles para ayudar a los buscadores como Google, Bing y otros a encontrar la información por medio de las etiquetas.[46]

5.2.16 **Peticiones HTTP**

---

[43] Javier J.Gutierrez ¿Qué es framework web? Tomado: http://www.lsi.us.es/~javierj/investigacion_ficheros/Framework.pdf

[44] JavaScrip Aprende a programar en el lenguaje Web, Luis Fernando O. Tomado de: https://books.google.es/books?hl=es&lr=&id=SqikDwAAQBAJ&oi=fnd&pg=PA4&dq=que+es+html%2Bcss%2BJs&ots=pz6hW_0kFA&sig=n7WDi8D05qM4zeoCwB4GFI4bSMM#v=onepage&q=que%20es%20html%2Bcss%2BJs&f=false

[45] 29 abril 2021 MDN Web Docs Tomado de: https://developer.mozilla.org/es/docs/Glossary/CSS

[46] Javier Flores Herrera,Codigofacilito Tomado de: https://codigofacilito.com/articulos/que-es-html

Definen un conjunto de métodos de petición para indicar la acción que se desea realizar para un recurso determinado. También son llamadas HTTP verbs[47]

GET: Solicita información.
HEAD: Pide una respuesta idéntica a la de una petición GET.
POST: Envía información
PUT: Actualiza los recursos en específico.
DELETE: Borra un recurso en específico.
CONNECT: Establece un túnel hacia el servidor identificado por el recurso.
OPTIONS: Opciones de comunicación para el recurso de destino.
TRACE: Prueba de bucle de retorno de mensaje a lo largo de la ruta al recurso de destino.
PATCH: Modificaciones parciales a un recurso.

- **5.3 MARCO JURÍDICO**

En Colombia el marco legal para la protección de delitos informáticos está en la Ley 1273 del 5 de enero de 2009. Según el Ministerio de las TIC, no existen normas que regulen específicamente las redes sociales.

La Ley en mención es una modificación del Código Penal, creando un nuevo bien jurídico; aquel que intenta protegernos de los delincuentes en la red.[48]

¿Dónde denunciar los delitos informáticos en Colombia?

En Colombia se puede informar sobre los delitos informáticos y denunciarlos a través de los sitios web:

- wwww.delitosinformaticos.gov.co

- www.policia.gov.co

Constitución Política de Colombia Artículo 61

El objetivo de la carta política, como norma superior, es indicar la responsabilidad que tiene el Estado colombiano para garantizar los derechos de los ciudadanos frente a la propiedad intelectual e indicar que, mediante las normas sustanciales y procesales, se puede proteger en cualquier tiempo siempre y cuando se atienda a las formalidades contenidas.

[47] MDN Web Docs 2021 Tomado de MDN Web Docs, https://developer.mozilla.org/es/docs/Web/HTTP/Methods

[48] Tomado de http://www.duende.com.co/blog/item/58-las-leyes-en-internet-y-los-delitos-informaticos#:~:text=En%20Colombia%20el%20marco%20legal,5%20de%20enero%20de%202009.&text=La%20Ley%20en%20menci%C3%B3n%20es,los%20delincuentes%20en%20la%20red.

Ley 23 de 1982 Sobre derechas de autor

Es una disposición sobre los derechos de autor. Mediante este instrumento se garantiza y regulan los derechos morales y patrimoniales que tienen los autores ante sus diversas creaciones, que bien pueden ser de carácter literario, científico o artístico, para proteger la invención humana. Asimismo, esta normativa determinó las limitaciones y excepciones a las que hay lugar, señalando las modalidades de tiempo y lugar. Fijó lo tendiente a su duración y reproducción sistémica.

Ley 527 de 1999 Comercio Electrónico

Es un instrumento que tiene como objetivo reglamentar el acceso y el uso de los mensajes de datos, del comercio electrónico y de las firmas digitales. Uno de sus principales fines fue la definición de conceptos claves para los efectos de ley que son tendientes a la materialización de una norma sustancial, de modo que, indica la forma de certificación y validez para que en efecto surta un reconocimiento.

Ley 599 de 2000 Código Penal

El código penal es una norma procesal que señala los tipos de prohibiciones y delitos, las diferentes sanciones para las conductas típicas, antijurídicas y culpables.
Esta norma indica la estructura del tipo penal en cuanto a la violación ilícita de comunicaciones en su artículo 192, indicando que aquel que actúe contrario a la ley mediante ciertas conductas configurativas ante una comunicación privada será constituido de una pena de prisión, la cual se divide en dos categorías tendientes a la divulgación o no divulgación de esta y del provecho obtenido, por ser un acceso abusivo a un sistema informático.
Así mismo, esta ley reguló en su artículo 269A el acceso abusivo a un sistema informático, indicando que aquel que sin autorización o fuera de lo acordado acceda a un sistema protegido o no, incurrirá en dos modalidades de sanción, la pena de prisión y la multa.

Ley 1273 de 2009

Es una modificación y adición que tuvo lugar en el código penal, cuyo objetivo fue la creación de un bien jurídico tutelable, el cual va encaminado a la protección de la información y los demás datos. Tiene como objeto la prevención y sanción de los delitos cometidos contra los sistemas o los realizados mediante el uso tecnologías.

Decreto 1727 de 2009
Es una decisión tomada por el Ministerio de Hacienda y Crédito Público decretada por el presidente de la República, con el fin de determinar la forma en la cual los operadores de los bancos de datos de información financiera, crediticia, comercial, de servicios y la proveniente de terceros países, deben presentar la información

de los titulares de la información. Tuvo como objeto la regulación para los efectos de la Ley 1266 de 2008 y se encaminó como norma sustancial, dado que fijó los requisitos y lo tendiente a los sectores financieros, los requerimientos de las entidades, y señaló la vigencia que tendría el decreto.

Decreto 2952 de 2010

Es una decisión tomada por el Ministerio de Hacienda y Crédito Público decretada por el Presidente de la República que tiene por objeto desarrollar el derecho constitucional que tienen los individuos de conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas. Mediante este decreto se regularon los artículos 12 y 13 de la Ley 1266/2008 tendientes a la permanencia de la información y el contenido de la información con el fin de garantizar la libertades y derechos de los ciudadanos.

Ley 1581 de 2012

La ley estatutaria constituye el marco general de la protección de los datos personales. Tiene como objetivo desarrollar el artículo 15 de carta política, dado que en él se garantiza el derecho a la intimidad personal y familiar, el derecho que tienen de conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas. Esta norma señala los principios rectores y lo concerniente con el tratamiento de datos y prohíbe la transferencia de datos personales de cualquier tipo.

Decreto Nacional 1377 de 2013

El objetivo de este decreto fue reglamentar parcialmente la Ley 1581 de 2012, de modo que indicó la autorización de recolección y el tratamiento de datos, fijó la forma en que debía desistirse de dicha autorización. De igual manera, señaló la política de tratamiento de datos, fijando las pautas, el contenido y la difusión de datos, indicó los derechos de los titulares, las transferencias y transmisiones internacionales de datos personales, y, la responsabilidad atribuida.

Decreto 886 de 2014

Tiene como objetivo reglamentar la información mínima que debe contener el Registro Nacional de Bases de Datos, creado por la Ley 1581 de 2012, así como los términos y condiciones bajo las cuales se deben inscribir en este los responsables del Tratamiento.

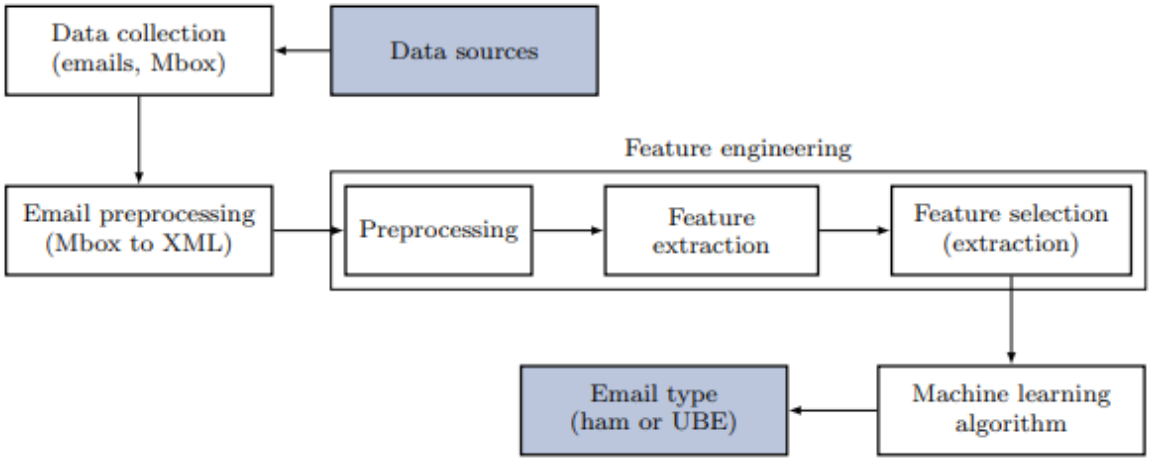
Convenio sobre Ciberdelincuencia 85 del Consejo de Europa – CCC (conocido como convenio sobre cibercriminalidad de Budapest)
Es una norma perteneciente al bloque de constitucionalidad a la que se adoptan los estados de forma facultativa. Su objetivo principal es la creación de normas internacionales que prevengan conductas delictivas de modo que se detecte, investigue y sancione conductas antijurídicas configurativas del delito cibernético.

6 ESTADO DEL ARTE

En el estado del arte se escogieron diferentes documentos publicados entre los años 2009 y 2020, Se evidenció que para la detección de ataques de phishing en correos electrónicos se han implementado diferentes técnicas y se han utilizado diferentes herramientas.

En el año 2020 Tushaar Gangavarapu,Jaidhar y Bhabesh Chanduka hacen una publicación acerca del filtrado de correo electrónico de spam y phishing. En este documento llamado "Aplicabilidad del aprendizaje automático en el filtrado de correo electrónico de spam y phishing: revisión y enfoques", explican las metodologías implementadas para la detección de correos electrónicos masivos no solicitados (UBE). Para este documento, el enfoque de selección de características basado en FI (usando RF), se clasificaron usando un clasificador de RF, y dio como resultado una precisión general del 98,4%.[49]

Figura 12. Descripción procedimiento empleado para extraer inferencias de los datos.



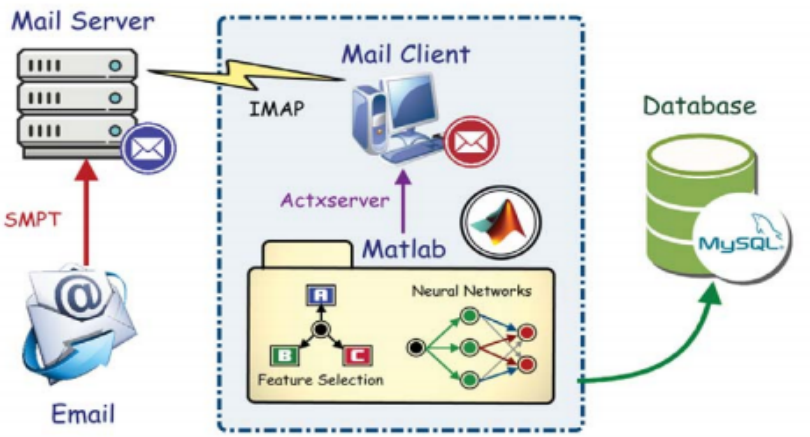
Tomado de: Applicability of Machine Learning in Spam and Phishing Email Filtering: Review and Approaches

En la Universidad de las Fuerzas Armadas ESPE en Ecuador, Diego Ona, Lenín Zapata, Walter Fuertes, Germán Rodríguez, Eduardo Benavides y Theofilos Toulkeridis presentaron en el año 2019 un documento llamado "Ataques de suplantación de identidad: detección y prevención de correos electrónicos infectados mediante métodos de aprendizaje automático" basándose en una herramienta (matlab) para la implementación del algoritmo, el cual detecta los ataques de phishing, desarrollando la metodología Agile Scrum. Además, se

[49] Tushaar Gangavarapu,Jaidhar y Bhabesh Chanduka.2020. Applicability of Machine Learning in Spam and Phishing Email Filtering: Review and Approaches Disponible en: https://tushaargvs.github.io/assets/publications/aire-2020-draft.pdf

basaron en la cámara de compensación colaborativa de datos e información sobre phishing en Internet PhishTank.[50]

Figura 13. Diagrama propuesto en el documento de la arquitectura



Tomado de: Detecting and Preventing Infected E-mails Using Machine Learning Methods

En el año 2014, en la Universidad de KwaZulu-Nata en Sudáfrica, Andronicus A. Akinyelu and Aderemi y O. Adewumi, en el artículo investigativo “Clasificación del correo electrónico de phishing mediante la técnica de aprendizaje automático de bosque aleatorio” se enfocaron en desarrollar un clasificador de correos electrónicos de phishing. A partir de un conjunto de datos que consta de 2000 correos electrónicos de phishing y ham, que fueron extraídos y utilizados por el algoritmo de aprendizaje automático de bosque aleatorio con una precisión de clasificación resultante del 99,7% y bajas tasas de falsos negativos (FN) y falsos positivos (FP).[51]

En la universidad de New Brunswicka, Ali Ghorbani, Huajie Zhang y Chair Rongxing Lu presentaron el documento “Un enfoque de detección de correo electrónico phishing utilizando técnicas de aprendizaje máquina”. Este proyecto lo denominaron Phishing Alerting System (PHAS), y tiene la capacidad de detectar y alertar todo tipo de correos electrónicos engañosos para ayudar a los usuarios en la toma de decisiones.[52]

En el año 2017 Srishti Rawal,Bhuvan Rawal,Aakhila Shaheen y Shubham Malik presentaron el documento “Detección de phishing en correos electrónicos mediante aprendizaje automático”. Para la clasificación de correos electrónicos

[50] Diego Ona,Lenín Zapata, Walter Fuertes, German Rodriguez,Eduardo Benavides y Theofilos Toulkeridis.2019. Phishing Attacks: Detecting and Preventing Infected E-mails Using Machine Learning Methods.
[51] Andronicus A. Akinyelu and Aderemi y O. Adewumi.2014. Classification of Phishing Email Using Random Forest Machine Learning Technique.
[52] Ali Ghorbani, Huajie Zhang y Chair Rongxing.2017. A PHISHING E-MAIL DETECTION APPROACH USING MACHINE LEARNING.

utilizaron el método de clasificación de SVM (Support Vector Machines) y el clasificador Random Forest**.** Con este método se precisó un máximo del 99. 87% en la clasificación.[53]

Shamal M. Firake, Pravin Soni, and B.B. Meshram del año 2011 en el documento "Herramienta para la prevención y detección de correos electrónicos de phishing Ataques", buscaban construir una herramienta que detecte correos maliciosos utilizando un conjunto de funciones de hipervínculo, de la lista negra y la lista blanca de URLs. con el fin de detectar ataques de phishing. Además, cuenta con una interfaz amigable con el usuario, para que este pueda manipularla. Este módulo lo desarrollan con tecnología Java[54].

En el año 2020 Suraj J Pai,Rakshitha Gokuldas, Rahul Kakkadan, Sourabh Hegde, Ms. Saritha Suvarna, estudiantes del departamento de ingeniería informática del Canara Engineering College, presentaron el "Analizador de sitios web de phishing para garantizar la banca electrónica y sitios web de comercio electrónico". El uso de Phishing Website Analyzer puede ayudar a los usuarios a predecir si el sitio web que utilizan es seguro mediante el uso de aprendizaje automático.[55]

Andronicus A. Akinyelu and Aderemi O. Adewumi del año 2014 presentaron el artículo "Clasificación de correos Phishing usando la técnica de Machine Learning Random Forest". En éste investigan sobre el uso de algoritmos de aprendizaje automático Random Forest en la clasificación de ataques de phishing. El objetivo fue desarrollar un clasificador de correo electrónico de phishing mejorado con una mayor precisión de predicción. Allí se concluye que la máquina forestal aleatoria utilizó una precisión de clasificación resultante del 99,7% y un nivel bajo de falsos negativos (FN) y falso tasas positivas (F).[56]

[53] Srishti Rawal,Bhuvan Rawal,Aakhila Shaheen y Shubham Malik.2017. Phishing Detection in E-mails using Machine Learning.
[54] Firake, S. M., Soni, P., & Meshram, B. B. (2011). Tool for Prevention and Detection of Phishing E-Mail Attacks.
[55] Suraj J Pai,Rakshitha Gokuldas, Rahul Kakkadan, Sourabh Hegde, Ms. Saritha Suvarna 2020. PHISHING WEBSITE ANALYZER TO SECURE E-BANKING AND E-COMMERCE WEBSITES
[56] Andronicus A. Akinyelu and Aderemi O. Adewumi,2014. Classification of Phishing Email Using Random Forest Machine Learning Technique.

7 METODOLOGÍA

7.1 METODOLOGÍA PROPUESTA

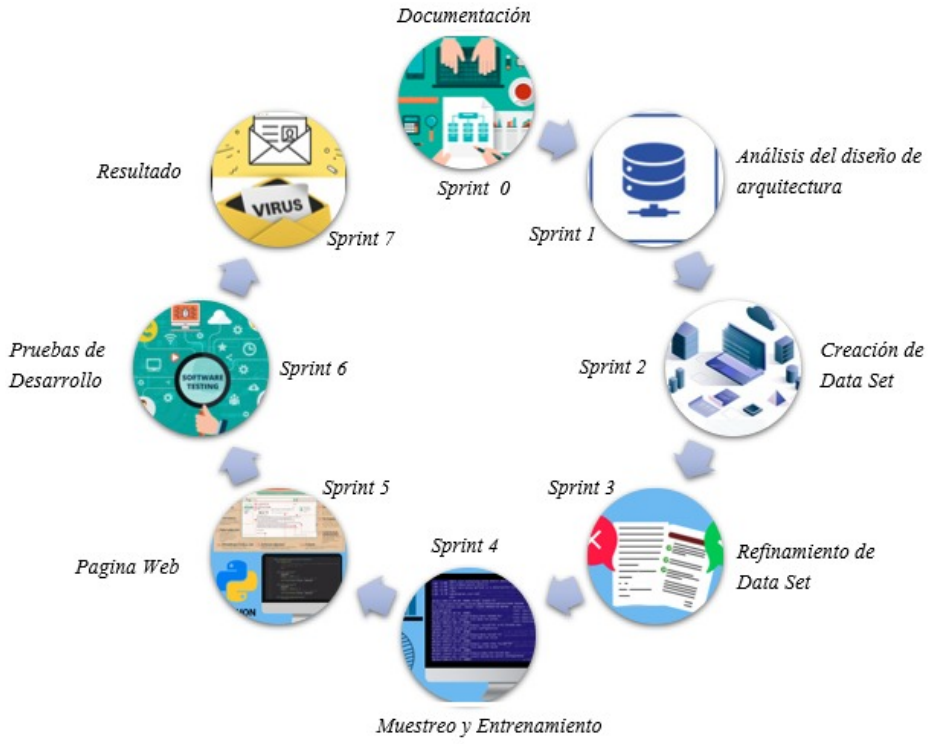
Para llevar a cabo el proyecto cuyo objetivo es la detección de Spear Phishing en correos electrónicos, se ha considerado pertinente implementar la metodología Scrum, ya que esta permite hacer un seguimiento de proyectos bajo un esquema de Sprints incrementales, buscando como resultado que cada avance sea presentado en línea con los objetivos finales del proyecto.

Estos Sprint mencionados anteriormente, se dan como completos una vez se realice la entrega de los compromisos adquiridos por los stakeholders o actores como lo son los: documentos, diseños, códigos, implementación, etc.

Dentro del desarrollo del proyecto se programan reuniones con los scrum team y estas se programarán un día en la semana con preferencia fin de semana. En estas reuniones se presentará un informe ejecutivo de los avances de cada Sprint, de igual manera, los inconvenientes y retroalimentación al proceso y se definen nuevos entregables con fechas siguientes de entregas.

A partir de lo anterior, se define que para este proyecto se contemplarán 8 Sprint como lo muestra la figura 7, para así establecer si el correo electrónico procesado contiene Spear Phishing.

Figura 14. Metodología para la Detección Spear Phishing

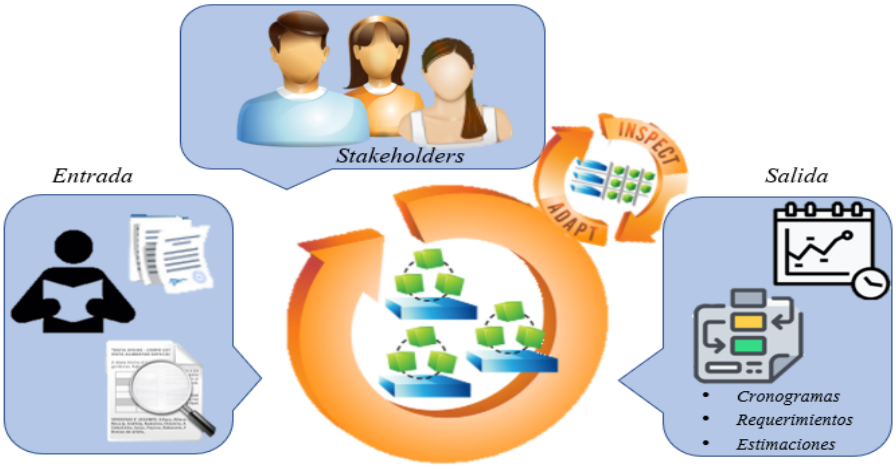


Tomado de: Autores

Sprint N° 0. Documentación

En este Sprint se reúnen los stakeholders y se busca que se definan los aspectos como funcionalidad, objetivos, riesgos, plazos de entrega, el componente de documentación el cual tendrá la asesoría con los expertos. En esta sesión se considera aplicar la técnica de lluvias de ideas. Como resultado de estas, se tiene el planteamiento de las estrategias.

Figura 15. Sprint N° 0. Documentación

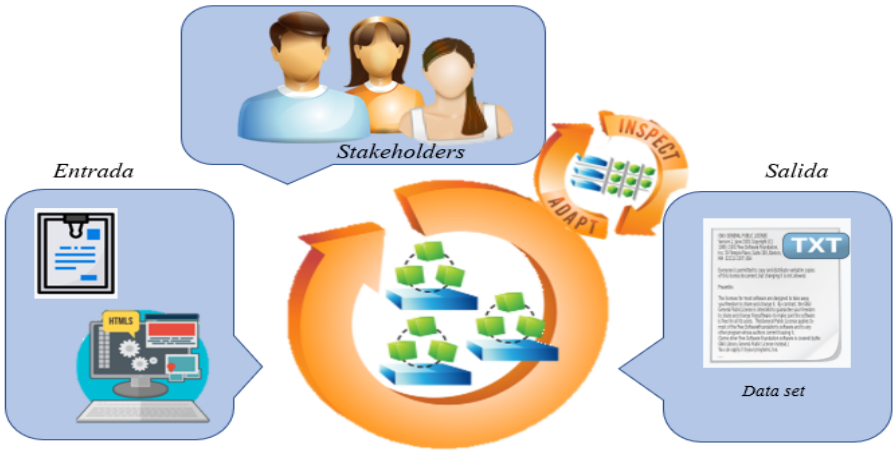


Tomado de: Autores

Sprint N° 1. Creación de Dataset

Bajo el mismo esquema de reunión, se realizará la búsqueda de diferentes datos como son correos electrónicos fraudulentos y no fraudulentos, los cuales sirven para desarrollar algoritmos que detectan si el correo caso estudio es phishing o no.

Figura 16. Sprint N° 1. Creación de Dataset

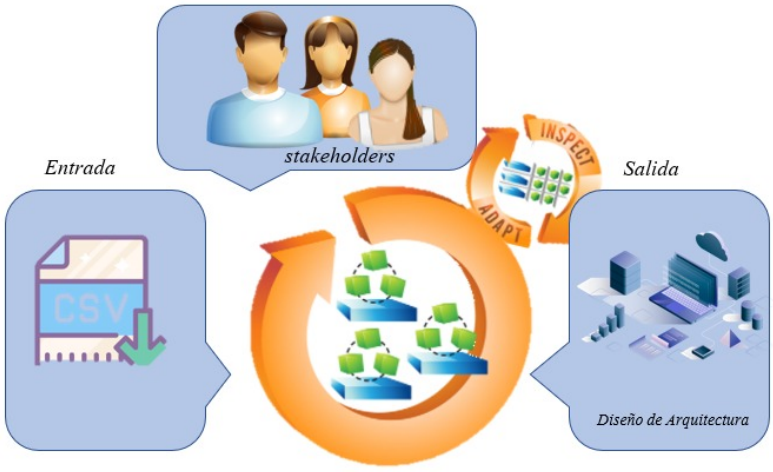


Tomado de: Autores

Sprint N° 2. Análisis del diseño de arquitectura

Una vez realizado el Dataset, se efectuará el análisis del diseño de arquitectura que se implementará en el prototipo propuesto.
Al finalizar el análisis se define el diseño de la arquitectura del diseño propuesto.

Figura 17. Sprint 2 análisis del diseño de arquitectura

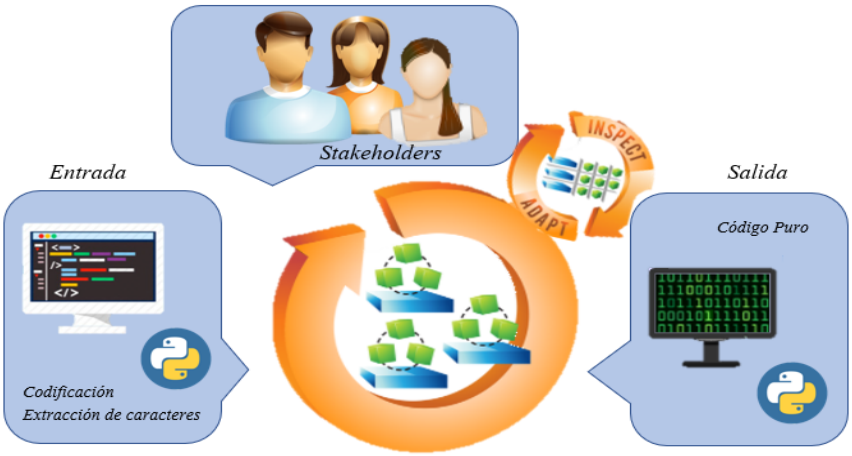


Tomado de: Autores

Sprint N° 3. Refinamiento del Dataset (NLP)

En este Sprint se inicia la fase de programación, la cual pretende procesar el Dataset, que, a su vez permite la extracción de caracteres o palabras no relevantes, iniciando así el Sprint del muestreo y entrenamiento del algoritmo.

Figura 18. Sprint N°3. inicialización Dataset

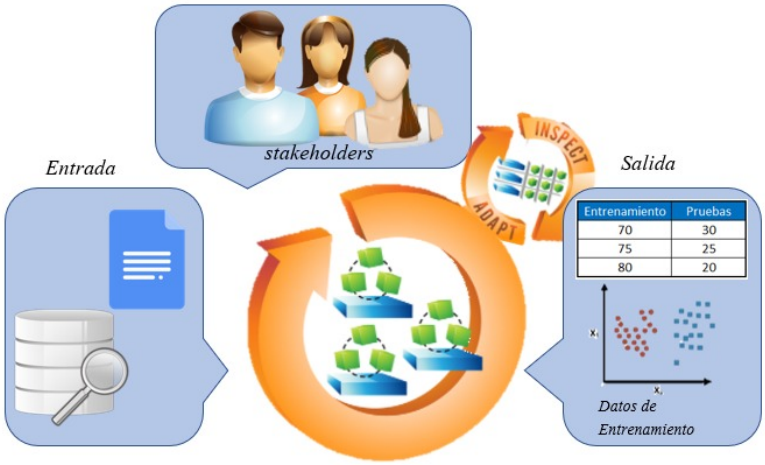


Tomado de: Autores

Sprint N° 4. Muestreo y entrenamiento del Algoritmo

Para llevar a cabo este Sprint se toman los datos y se dividen en dos conjuntos con el fin de obtener la información del entrenamiento y pruebas del algoritmo. Con la técnica validación cruzada o en inglés cross validation, se evaluarán los resultados para el análisis estadístico para la partición de los datos.

Figura 19. Sprint N° 4. Muestreo del Algoritmo

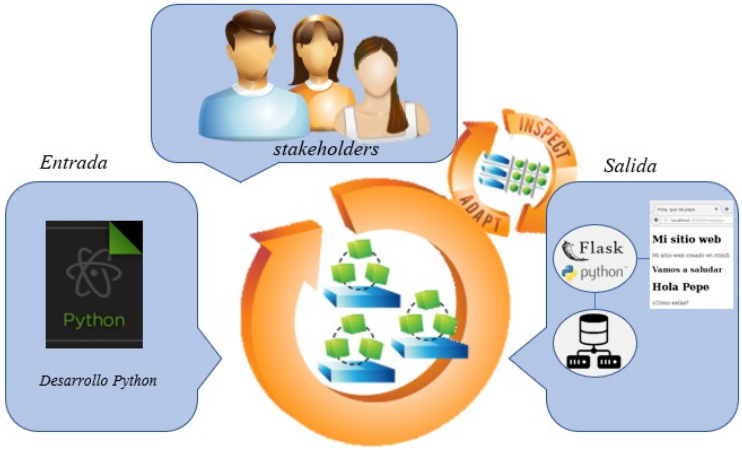


Tomado de: Autores

Sprint N° 5. Página web

En este Sprint se desarrollarán los mockups del sitio web. Adicional, en las reuniones con los stakeholders, se decidirá qué framework usar para el desarrollo de la página, buscando la compatibilidad con el lenguaje de programación Python, ya que este leguaje permite la integración del Dataset y la página, la cual realizará la clasificación de si el correo es fraudulento.

Figura 20. Sprint N° 5. Página web

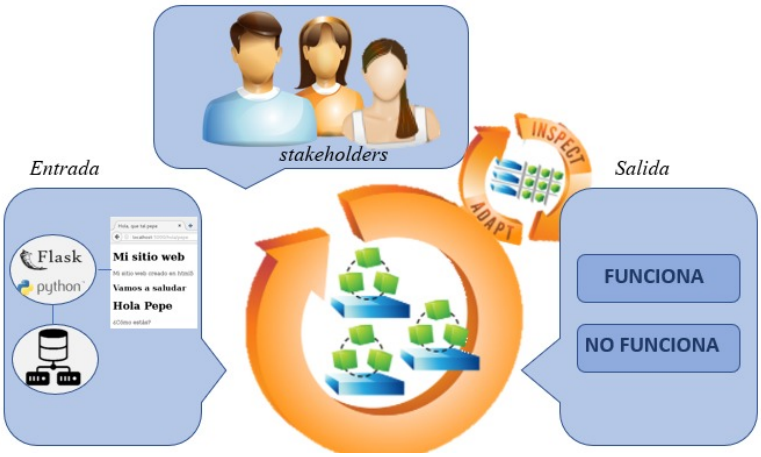


Tomado de: Autores

Sprint N° 6. Pruebas del desarrollo

En el Sprint 6 el prototipo se empleará de manera experimental para asegurarse de que este no tenga fallas, es decir, que funcione de acuerdo con las especificaciones. En este Sprint se realizarán una serie de pruebas con conjunto de datos reales. Para esto, se harán reuniones con los stakeholders, donde cada uno evidenciará el proceso de pruebas.

Figura 21. Sprint N° 6. Pruebas del desarrollo

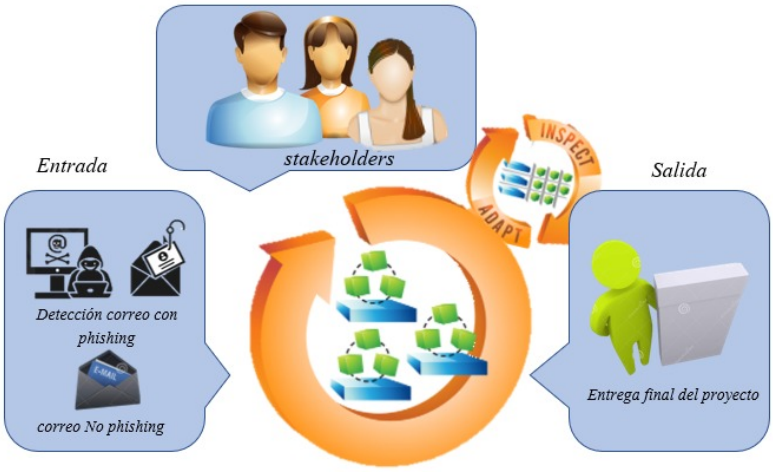


Tomado de: Autores

Sprint N° 7. Resultados

En este último Sprint, y una vez verificado el entorno de integración, los equipos realizarán una reunión para la entrega de los paquetes del proyecto, en la cual se establecerá el documento final y la correcta funcionalidad del prototipo propuesto. En este documento se mencionarán a los stakeholders que participaron en el proyecto de detección de correos electrónicos Spear Phishing.

Figura 22. Sprint N° 7. Resultados



Tomado de: Autores

8 DESARROLLO DE LA PROPUESTA

8.1 CREACIÓN DEL DATASET

Para la implementación y prueba del prototipo empleado, se usaron un conjunto de datos disponibles públicamente, en las plataformas gratuitas llamadas Kaggle y Github.

Kaggle es una comunidad en línea, donde se relacionan científicos de datos y profesionales en aprendizaje automático. En esta plataforma se obtuvieron 3 conjuntos de datos llamados:

Spam or not Spam Dataset[57]
Spam Mails Dataset[58]
Fraudulent E-mail Corpus[59]

Github es una plataforma que permite que todos los usuarios puedan aportar códigos de proyectos. Además, este repositorio tiene una gran ventaja, ya que su plataforma deja descargar y revisar todos los proyectos que estén abiertos. Igualmente se puede configurar si el usuario prefiere que su código no esté disponible. En esta plataforma se adquirieron cinco conjuntos de datos llamados:

Phishingdata-Analysis[60]
Machine-Learning-with-R-Datasets[61]
spam.csv[62]
Phishtank_Dataset.csv[63]
sms_spam.csv[64]

Adicionalmente se recopilaron correos phishing de personas naturales y empresas logrando así recolectar un total de 3.725 correos los cuales fueron etiquetados.

8.2 ANÁLISIS DEL DISEÑOS DE ARQUITECTURA

Arquitectura de Software

[57] Hakan Ozler https://www.kaggle.com/ozlerhakan/spam-or-not-spam-dataset
[58] Venkatesh https://www.kaggle.com/venky73/spam-mails-dataset
[59] Rachael Tatman
https://www.kaggle.com/rtatman/fraudulent-email-corpus?select=fradulent_emails.txt
[60]TanusreeSharma
https://github.com/TanusreeSharma/phishingdata-Analysis/tree/master/1st%20data
[61] Stedy https://github.com/stedy/Machine-Learning-with-R-datasets/blob/master/sms_spam.csv
[62] Mmerce https://github.com/bigmlcom/python/blob/master/data/spam.csv
[63] CbEkanayake 1209973
https://github.com/1209973/Phishing-Detection/blob/master/Phishtank_dataset.csv
[64] Zach Stednick
https://github.com/stedy/Machine-Learning-with-R-datasets/blob/master/sms_spam.csv

Cuando se habla de arquitectura se piensa que es la forma en la que los diferentes componentes del edificio se integran para formar un todo unido. Puesto que la arquitectura es la forma en que el edificio encaja en su entorno y con los otros edificios de su vecindad.

Con esto se puede decir que la arquitectura de software se basa en un conjunto de estructuras, propiedades y relaciones que conforman el "edificio" y esto permitirá dar soporte a la solución software que se va a desarrollar. Su desarrollo es de vital importancia básicamente por tres razones claves:

1. Facilita la comunicación entre todas las partes interesadas en el desarrollo del software.
2. Permite tomar decisiones tempranas y evitar un impacto mayor en una etapa posterior del desarrollo del proyecto.
3. Permite comprender fácilmente la estructura y el flujo de trabajo de sus componentes.[65]

Es por esto que elegir una arquitectura adecuada ayudará a proporcionar la funcionalidad deseada y los atributos de calidad, ya que es importante entender las diferentes arquitecturas antes de aplicarlas a nuestro diseño. Debido a que la mayoría de los sistemas son distintos entre sí, existen similitudes entre las arquitecturas. Estas suelen seguir lo que se conoce como patrón o estilo arquitectónico que captan la esencia de una arquitectura que se usa en diferentes sistemas de software.[54]

Algunos de los patrones arquitectónicos son:

8.2.1 **Patrón de Capas**

El patrón de capas es útil para estructurar programas que se pueden descomponer en grupos de subtareas. Cada capa se encuentra en un nivel particular de abstracción y estas proporcionan servicios a la siguiente capa superior.

Las 4 capas más comúnmente encontradas son:
- Capa de presentación (también conocida como capa UI)
- Capa de aplicación (también conocida como capa de servicio)
- Capa de lógica de negocios (también conocida como capa de dominio)
- Capa de acceso a datos (también conocida como capa de persistencia)

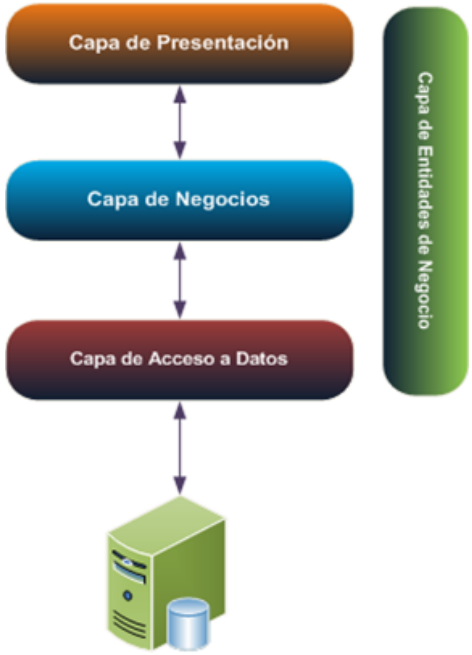
Los usos más comunes para este patrón de capas se dan en las aplicaciones de escritorio generales y aplicaciones web de comercio electrónico.[66]

---

[65] Segura Ariel Alejandro, Arquitectura de Software de Referencia para Objetos Inteligentes en Internet de las Cosas

[66] Septiembre,2018 Wilber Ccori huaman, Tomado de: https://medium.com/@maniakhitoccori/los-10-patrones-comunes-de-arquitectura-de-software-d8b9047edf0b

Figura 23. Patrón de Arquitectura de capas



Tomado de: https://analisisdesistemas1.files.wordpress.com/2015/05/arquitecturacapas.png

8.2.2 **Patrón cliente-servidor**

Este patrón se refiere a un modelo de comunicación que vincula a varios dispositivos informáticos a través de una red. El cliente es el que envía una o varias peticiones al servidor a través de una red y este se encarga de satisfacer dichos requerimientos.

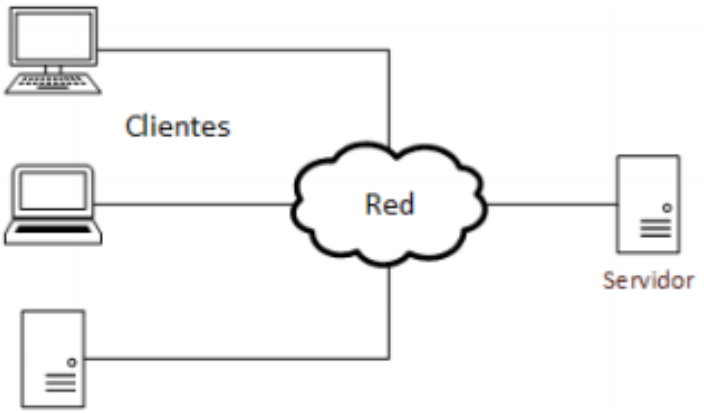
Los principales componentes de este modelo son:[12]

- Conjunto de servidores que ofrecen servicios a componentes. Como servidores web o de impresión.
- Conjunto de clientes que solicitan los servicios que ofrecen los servidores. Como navegadores web.
- Una red que permite a los clientes acceder a estos servicios. Como, por ejemplo, Internet.

Este patrón de cliente servidor es utilizado cuando el software necesita servir a varios clientes, como por ejemplo aplicaciones web, procesos de negocio que necesitan ser utilizados a lo largo de una organización o se están desarrollando servicios para que sean consumidos por otras aplicaciones.[67]

[67] Septiembre,2018 Wilber Ccori huaman,Tomado de: https://medium.com/@maniakhitoccori/los-10-patrones-comunes-de-arquitectura-de-software-d8b9047edf0b

Figura 24. Patrón cliente Servidor



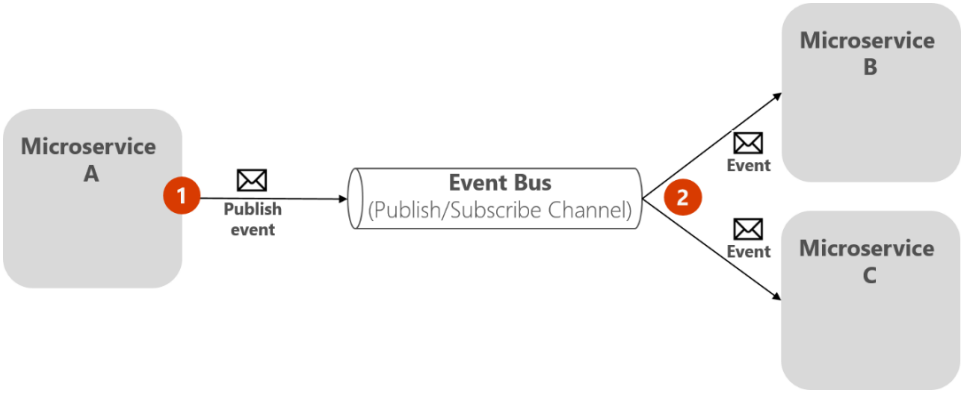
Tomado de: Arquitectura de Software de Referencia para Objetos Inteligentes en Internet de las Cosas.

8.2.3 **Patrón de bus de evento**

En este patrón trata principalmente de eventos y en él se observan cuatro componentes principales: la fuente de evento escucha de evento, canal de evento y bus de evento.[13] Permite a objetos suscribirse a ciertos eventos del bus, así que cuando un evento es publicado en el bus, se propaga a cualquier suscriptor interesado, debido a que se centra en la generación y entrega de notificaciones.
El bus de eventos también puede ser diseñado como una interfaz a una API necesaria para suscribirse a eventos, cancelar las suscripciones y publicar eventos. Puede tener una o más implementaciones basadas en cualquier comunicación de mensajería o entre procesos, como una cola de mensajes o un bus de servicio que admita la comunicación asincrónica y un modelo de publicación/suscripción[68]. Se utiliza para desarrollo de Android y servicios de notificación.

Figura 25. Patrón Bus de evento

[68] 2021 Enero, Nish Anil Implementación de comunicación basada en eventos entre microservicios (eventos de integración) Tomado de: https://docs.microsoft.com/es-es/dotnet/architecture/microservices/multi-container-microservice-net-applications/integration-event-based-microservice-communications.



Tomado de:
https://docs.microsoft.com/es-es/dotnet/architecture/microservices/multi-container-microservice-net-applications/integration-event-based-microservice-communications

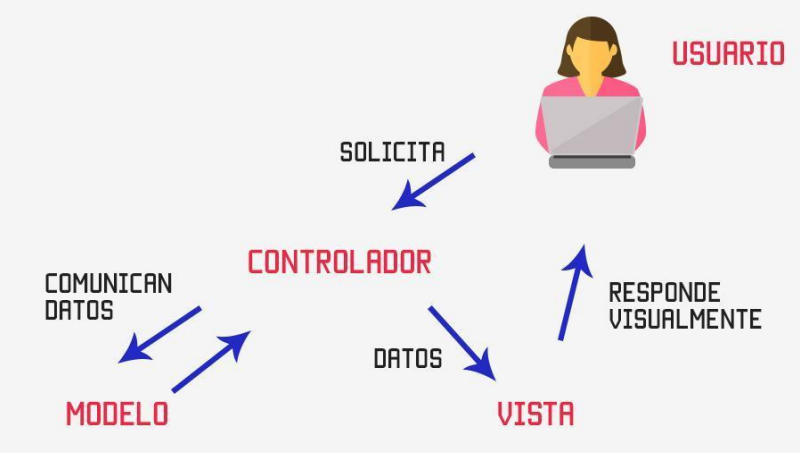
8.2.4 **Modelo-vista-controlador (MVC)**

También es conocido como patrón de arquitectura de tres capas, el cual separa presentación e interacción de los datos del sistema. El MVC se compone del modelo que maneja los datos del sistema y sus operaciones, la vista define y gestiona como se presentan esos datos al usuario y el controlador dirige la interacción del usuario con la interfaz de usuario mediante el manejo de botones[13]. Una de las ventajas de MVC es que permite separar los componentes de las aplicaciones dependiendo de la responsabilidad que tienen. Ello permite que cuando se desea realizar un cambio en alguna parte del código, este no afecte otra parte de la aplicación. Por ejemplo, si se modifica la base de datos, sólo se deberá modificar el modelo que es el encargado de los datos y el resto de la aplicación debería permanecer intacta. Esto respeta el principio de la responsabilidad única.[69]

Este tipo de modelo es utilizado para aplicaciones Web en los principales lenguajes de programación y en Marcos web como Django, flask y Rails.[70]

[69]Uriel Hernández, Codigofacilito MVC (Model, View, Controller) Explicado. Tomado de: https://codigofacilito.com/articulos/mvc-model-view-controller-explicado
[70]2018 Septiembre Wilber Ccori huaman Tomado de: https://medium.com/@maniakhitoccori/los-10-patrones-comunes-de-arquitectura-de-software-d8b9047edf0b

Figura 26. Patrón MVC



Tomado de: https://codigofacilito.com/articulos/mvc-model-view-controller-

Según la información recopilada de los tipos de arquitectura software, la que se optó para implementar en el desarrollo del prototipo propuesto es el MVC, debido a que este contiene ciertas particularidades como:

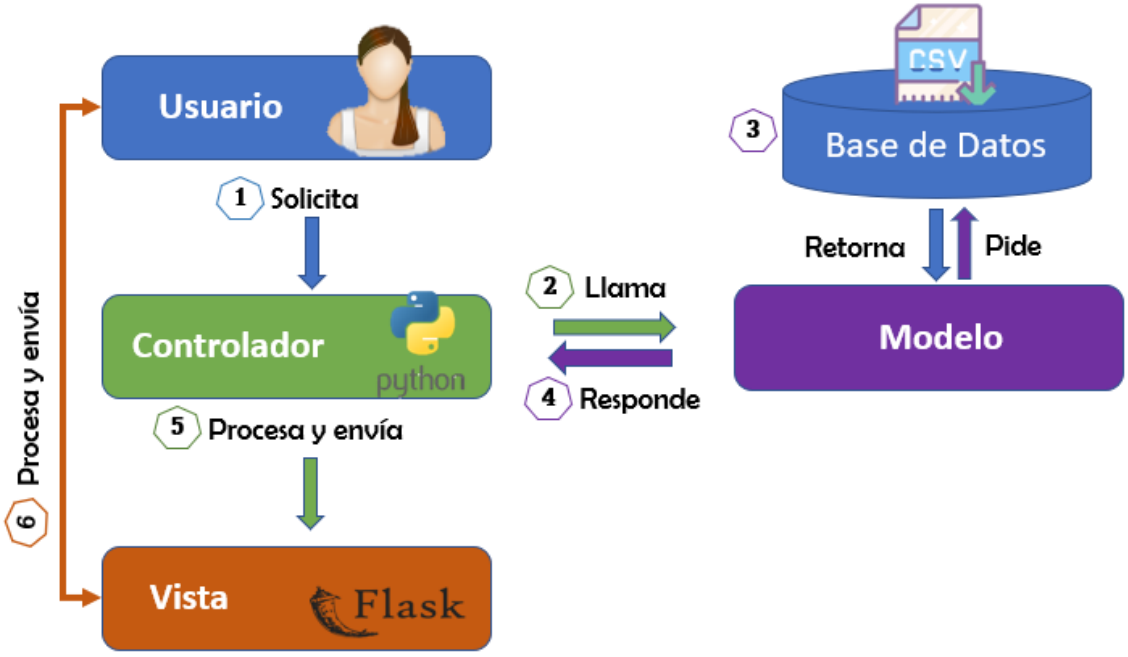
Proceso de desarrollo más rápido: ya que se puede trabajar en paralelo. Al utilizar este modelo, se puede desarrollar por un lado la vista y por otro lado se puede trabajar en el controlador. Con esto se puede reducir el tiempo de desarrollo al contar con varios colaboradores.
La modificación no afecta a todo el modelo: Cualquier cambio que se realice en el modelo no afectará a toda la arquitectura propuesta en la aplicación, porque la parte del modelo no es dependiente de algún otro componente como las vistas.
Fomentan el cambio: Las pruebas unitarias facilitan que el programador cambie el código para mejorar su estructura. Esto permite que al hacer cualquier tipo de cambio no afecte los ya construidos.
Facilidad de documentar el código: Las propias pruebas son documentación del código, puesto que ahí se puede ver cómo utilizarlo.
Separación de la interfaz y la implementación: esto es porque la única interacción entre las unidades que se están probando y los casos de prueba son las interfaces que las unen, esto quiere decir que en cualquier momento se pueden cambiar los dos si ningún efecto.
Los errores están más acotados y son más fáciles de localizar: Dado que tenemos pruebas unitarias que pueden desenmascararlos.

Por último, se decidió que la arquitectura MVC es la que más se ajusta al desarrollo propuesto dado a su diseño simplificado y robusto.

Ciclo de vida del MVC

El usuario realiza una petición.
El controlador captura la petición del usuario.
El controlador llama al modelo.
El modelo interactúa con la base de datos, y retorna la información al controlador.
El controlador recibe la información y la envía a la vista.
La vista procesa la información recibida y la entrega de una manera visualmente entendible al usuario.

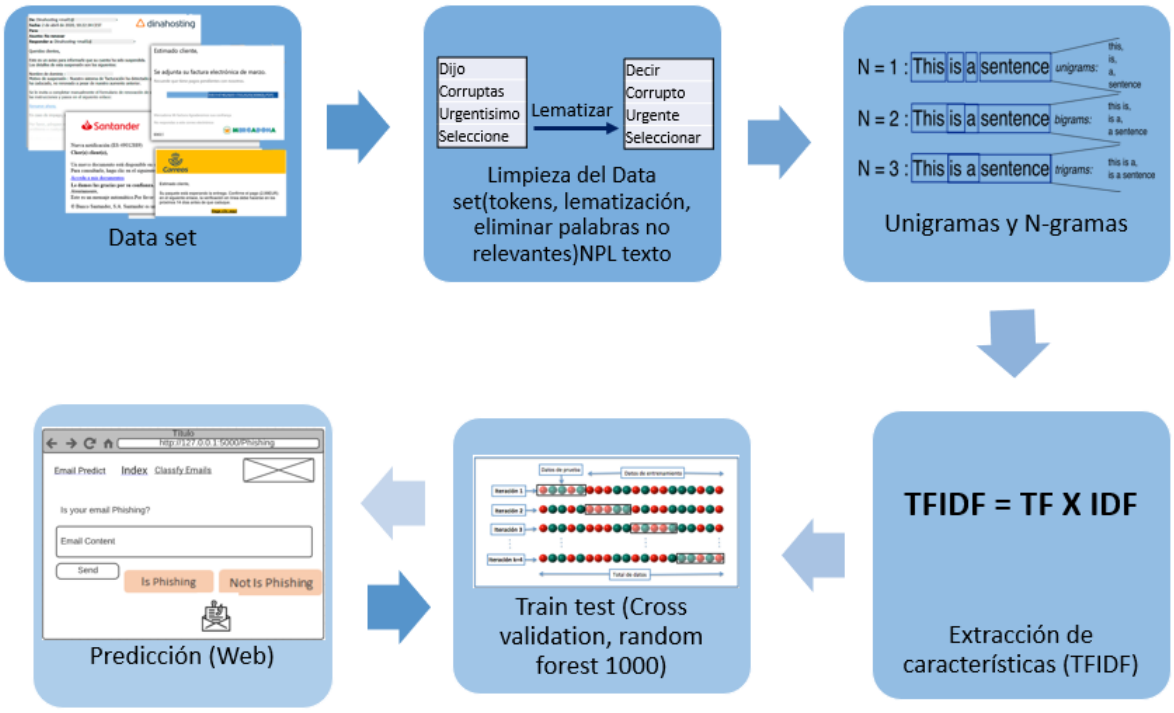
Figura 27. Diseño MVC del Prototipo Propuesto



Tomado de: Autores

Inicio de la fase de Programación

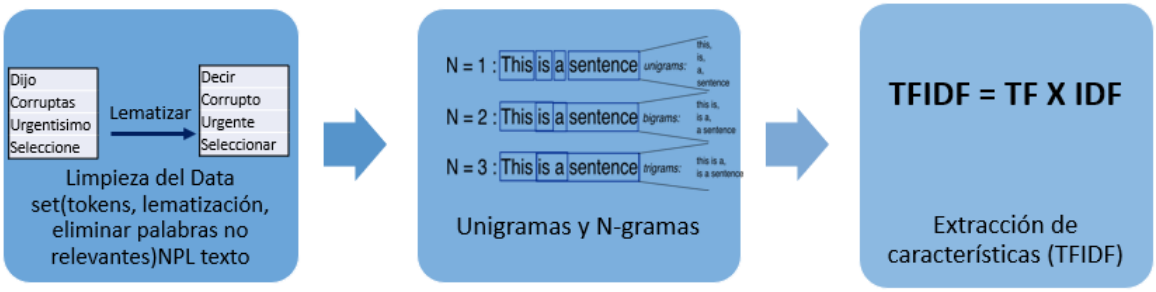
Figura 28. Fase de Programación



Tomado de: Autores

8.2.5 **Refinamiento del Dataset**

Figura 29. Refinamiento del Dataset



Tomado de: Autores

Una vez realizado el etiquetado del Dataset se procede a realizar el refinamiento o limpieza de este. Con el lenguaje de programación Python se procede a cargar los datos. Posteriormente se utilizó la técnica del procesamiento de lenguaje natural NPL a través de la lematización, Tokenizacion, eliminación de StopWord (conectores), eliminación de caracteres especiales y emoticones.

La Lematización es una de las mayores técnicas para la detección de Phishing debido a que se emplea, cuando se observan repeticiones en los términos (palabras) que contienen los correos electrónicos. La existencia de palabras derivadas de otras puede alterar los resultados de los cálculos. En efecto, resulta deseable reducir todos los vocablos derivados de una raíz común a un único término que es lo que se conoce como (lematización). Un proceso de extracción de raíces léxicas (stemming) podría implicar una mejora en los resultados obtenidos. Esto conlleva la eliminación de los prefijos y sufijos de cada término. En inglés, la ausencia de género en las palabras y otras características morfológicas del lenguaje simplifican esta tarea. [71]

En la figura 30 se procede a la eliminación de caracteres numéricos a través de un arreglo llamado *remove (list).*

Figura 30. Eliminación caracteres Numéricos

```
def remove(list):
    pattern = '[0-9]'
    list = [re.sub(pattern, '', i) for i in list]
    return list
```

Tomado de: Autores

Según se observa en la figura 30 se elige la táctica de los n-gramas, los cuales buscan las posibles combinaciones para así llegar a la palabra raíz.

Los n-gramas son medidas de asociación entre pares de términos, se calculan basadas en bigramas únicos compartidos (heurístico). Una vez encontrados los diagramas únicos se calcula la medida de similitud que se determina para todos los pares de términos de la base, formándose la matriz de similitu[72]

Ecuación 7. Fórmula de N-gramas

$$S = \frac{2C}{A+B}$$

A y B =Número de Diagramas únicos en el primer y segundo término
C = Número de Diagramas únicos compartidos por A y B

En este paso se procede con la limpia el Dataset (remover todos los caracteres especiales, enlace, emoticones, adicional convierte todo el texto en minúscula, realiza la lematización, stopword.

[71] 2007. José R. Méndez, Florentino Fdez-Riverola, Fernando Díaz, Juan M. Corchado. Sistemas inteligentes para la detección y filtrado de correo spam: una revisión
[72] Tomado de Universidad de Costa Rica facultad de ingeniería, Profesora M.Sc. Kryscia Ramírez Benavide, 12-11-2009, https://sites.google.com/site/jcorderoa41667/

Figura 31. n-gramas

```
def generate_ngrams(s, n):
    # Convert to lowercases
    s = s.lower()
    # Romper la oración en el tokens
    tokens_first = re.compile(r'('+'|'.join(regex_str)+')', re.VERBOSE | re.IGNORECASE )
    tokens = tokens_first.findall(s)

    # eliminar tokens vacíos, links, menciones, stopwords y hastags
    tokens = [token for token in tokens if (token != "" or token != '' )
              and (token not in stop)
              and (not token.startswith(('#', '@', 'https:', 'http:','<')))]
    #lematizar
    tokens = [stemmer.stem(token) for token in tokens]
    tokens = remove(tokens)
    tokens = [token for token in tokens if (token != "" or token != ''  )
              and (token not in stop)
              and (not token.startswith(('#', '@', 'https:', 'http:')))]
#     print(tokens)
    # funcion zip genera n-grams
    # Concatena los tokens en los ngrams y los retorna
    ngrams = zip(*[tokens[i:] for i in range(n)])
    return [" ".join(ngram) for ngram in ngrams]
```

Tomado de: Autores

Después de obtener el Dataset depurado (sin caracteres especiales, números, ni emoticones) se crea un arreglo llamado bigBagOfWords el cual guarda este nuevo conjunto de datos.

Figura 32.Arreglo BigBagOfWords

```
bigBagOfWords = []
unique = []
# bigTf = {}
# bigTfidf= []
##########Preprocesamiento##########
n=1
count = 0
count2 = 0
for email in emails:
    bagOfWords = generate_ngrams(str(email), n)
    unique = set(unique).union(set(bagOfWords))
    bigBagOfWords.append(bagOfWords)
```

Tomado de: Autores

En la Figura 26 con el arreglo bigBagOfWords se procede a realizar el cálculo de TF y IDF para luego extraer las características de dicho texto.

Figura 33. TD - IDF

```
def computeTF(wordDict, bagOfWords):
    tfDict = {}
    bagOfWordsCount = len(bagOfWords)
    for word, count in wordDict.items():
        if bagOfWordsCount != 0:
            tfDict[word] = count / float(bagOfWordsCount)
        else:
            tfDict[word] = 0

    return tfDict
```

```
def computeIDF(documents):
    import math
    N = len(documents)
    idfDict = dict.fromkeys(documents[0].keys(), 0)

    for document in documents:
        for word, val in document.items():

            if val > 0:
                idfDict[word] += 1
    for word, val in idfDict.items():
        if val != 0:
            idfDict[word] = math.log(N / float(val))
        else:
            idfDict[word] = 0
    return idfDict
```

Tomado de: Autores

Después de Calcular el TF y el IDF se utiliza la fórmula de TFIDF para realizar la extracción de características

$$TF - IDF = TF * IDF$$

Figura 34. Cálculo del TFIDF

```
def computeTFIDF(tfBagOfWords, idfs):
    tfidf = {}
    for word, val in tfBagOfWords.items():
        tfidf[word] = val * idfs[word]
    return tfidf
```

Tomado de: Autores

8.3 MUESTREO Y ENTRENAMIENTO

En este Sprint se utiliza el algoritmo Random Forest para realizar la clasificación, donde se seleccionan *k* columnas de las *m* totales que tienen los datos, en el que se crearon 1000 árboles de decisión, donde se pide que haga una misma

clasificación. Posteriormente se utilizó Cross-Validation, en el cual se dividieron los datos en varios conjuntos y cada conjunto se dividió en subconjuntos que representarán los datos de entrenamiento y prueba. Para esto se aplicaron cinco iteraciones, donde cada una de ellas contará con cuatro procesadores, como se observa en la figura 35.

Figura 35. Desarrollo Muestreo y Entrenamiento

```
In [3]: import numpy as np
        import pandas as pd
        from sklearn.model_selection import train_test_split
        from sklearn.preprocessing import StandardScaler
        from sklearn.ensemble import RandomForestClassifier
        from sklearn import metrics
        from sklearn.metrics import confusion_matrix
        from sklearn.externals import joblib
        from sklearn.metrics import classification_report, confusion_matrix
        from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
        from sklearn import preprocessing
        from sklearn.preprocessing import MinMaxScaler
        from sklearn.model_selection import cross_val_predict
        from sklearn.metrics import precision_score
        from sklearn.utils import shuffle
        dataset = pd.read_csv("Features unigrama.csv", sep=",", low_memory=False)
        shuffle(dataset)
        df = dataset.fillna(0)
        sc = StandardScaler()
        #Create a RandomForest Classifier
        clf=RandomForestClassifier(n_estimators=1000,bootstrap = True,oob_score=True, n_jobs=-1)
        y = df.label
        X = df.drop('label', axis=1)
        scaler = MinMaxScaler(feature_range=(0, 1))
        X = scaler.fit_transform(X)
        y_pred = cross_val_predict(clf, X, y, cv=5, n_jobs=4)
```

Tomado de: Autores

Luego de realizar todo el proceso de entrenamiento y prueba se empleó la Matriz de confusión para evaluar el modelo de aprendizaje automático basado en métricas, ya que el prototipo propuesto es de clasificación, como se muestra en la siguiente Figura 36.

Figura 36. Desarrollo Matriz de confusión

```
In [6]: print('Matriz de Confusión:')
        plot_confusion_matrix(cm = confusion_matrix(y,y_pred), target_names=['No Fraude', 'Fraude'],
                              title='Matriz de confusion', cmap=None, normalize=False)
        #Calculo la precisión del modelo
        precision = precision_score(y, y_pred,average='weighted')
        print('Precisión del modelo:')
        print(precision)

        df2 = pd.DataFrame(classification_report(y, y_pred,output_dict=True)).T
        df2
```

Tomado de: Autores

8.4 PAGINA WEB

En este Sprint se da inicio con el desarrollo de los mockups, tanto para el home de la página web como para el ingreso del texto del correo electrónico a revisar, como se observa en las siguientes figuras.

Figura 37. Mockups Index

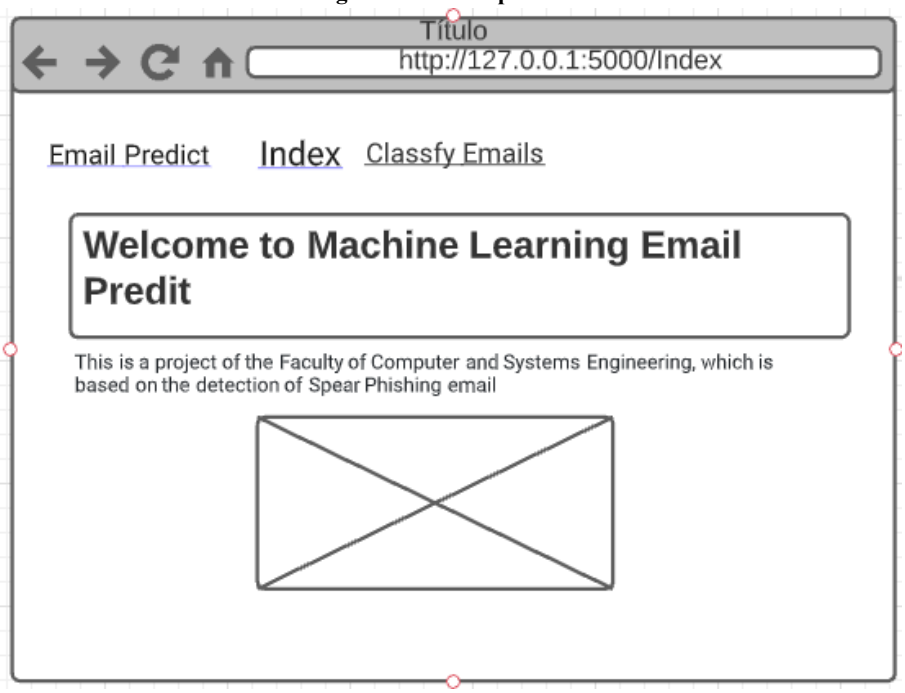
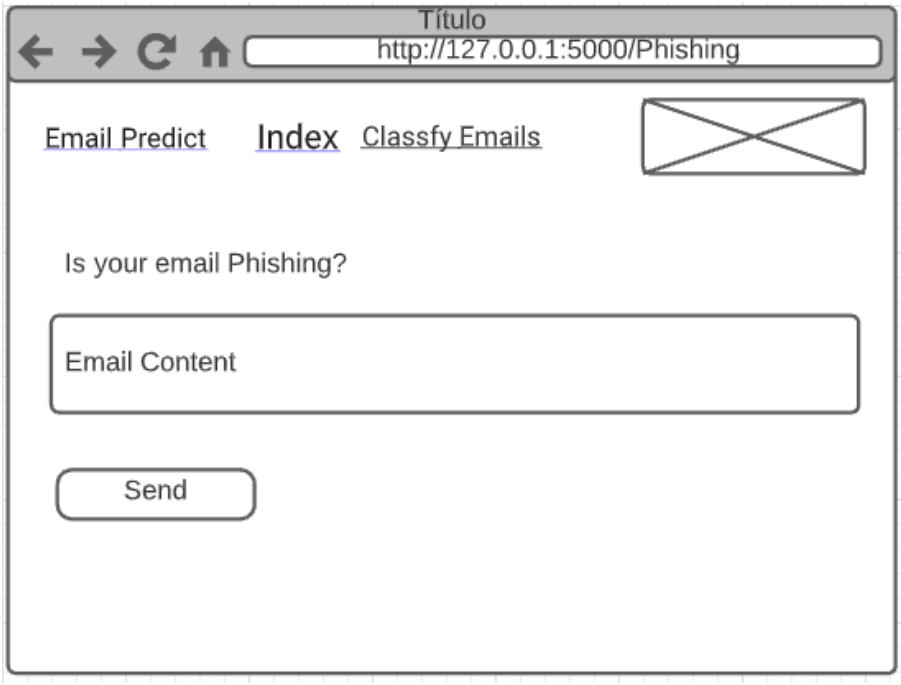


Figura 38. Mockups Phishing

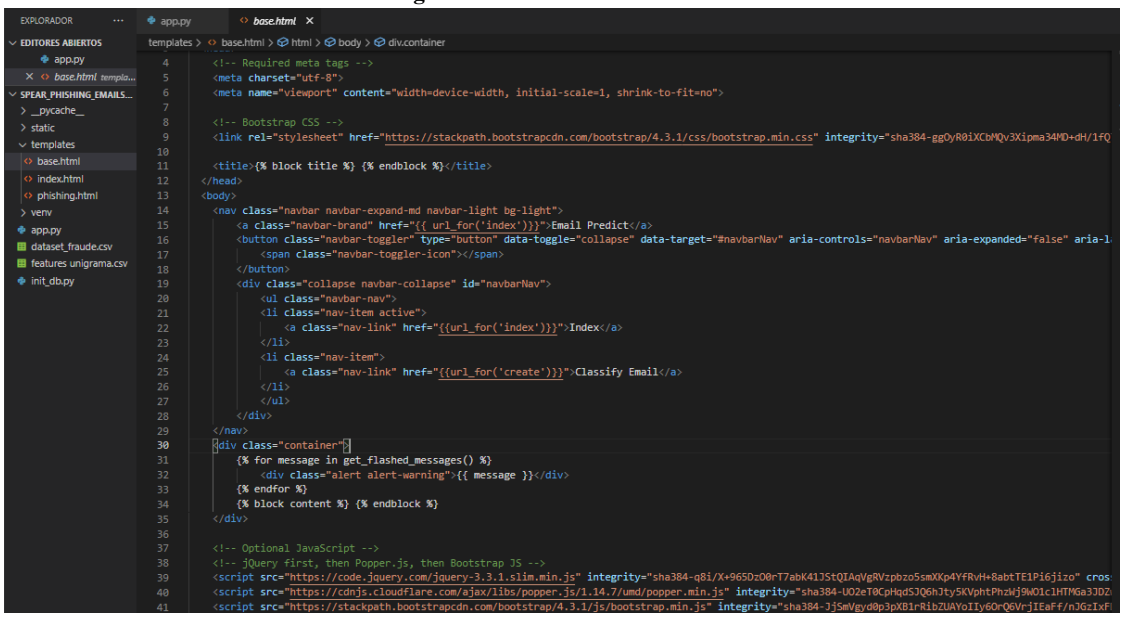


Tomado de: Autores

Luego de realizar estos diseños, se implementó la herramienta de desarrollo web *Visual Studio Code,* debido a que se puede usar con distintos lenguajes, que hacen que el trabajo con el software sea más agradable a la vista.

Se realizó la parte del *front end* con lenguaje HTML

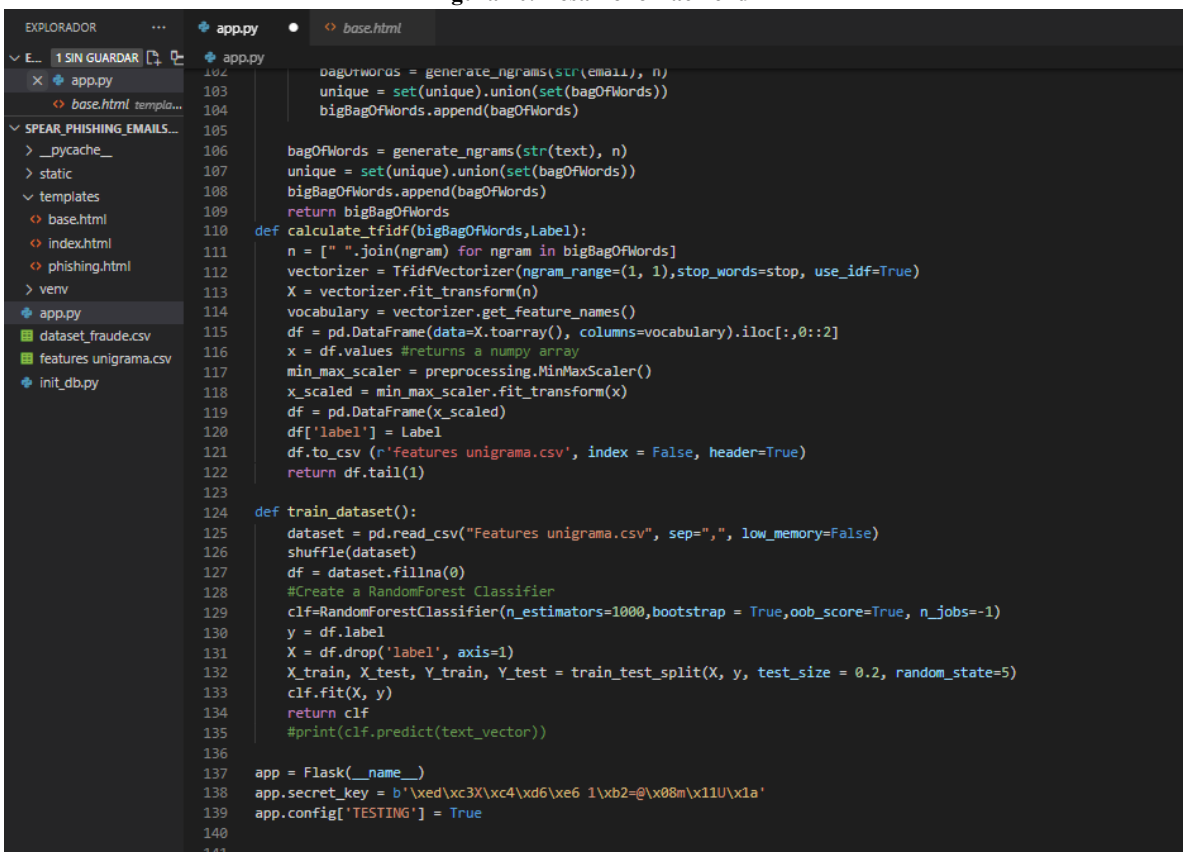
Figura 39. Desarrollo Front - end



Tomado de: Autores

Luego se acoplará lo realizado en Python como el *Back end*, del prototipo que tiene como nombre *app.py*

Figura 40. Desarrollo Back end



Tomado de: Autores

8.5 PRUEBAS DE DESARROLLO.

Para el desarrollo de este sprint se da inició con pruebas de humo (Smoke Tests) y luego con pruebas de funcionales (System Tests), las cuales validaron el correcto funcionamiento del prototipo desarrollado.

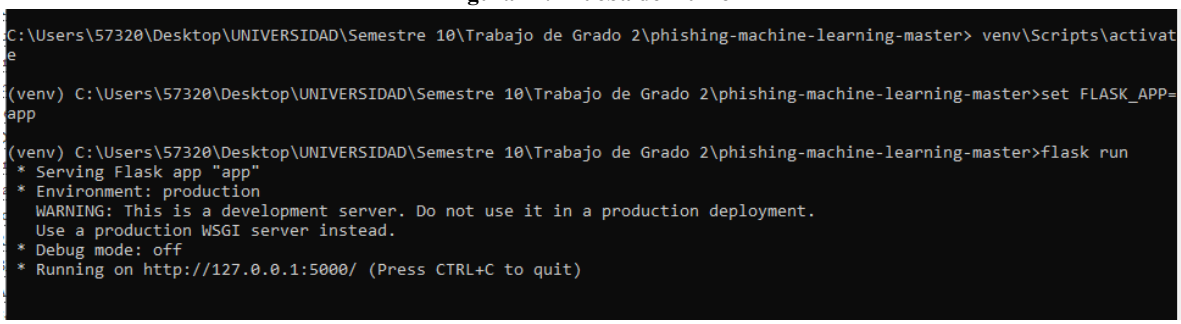
Las pruebas de humo son pruebas que verifican la funcionalidad básica de una aplicación. Su principal ventaja es ser una prueba rápida de ejecutar, con el objetivo de asegurar las características principales del sistema

Este tipo de prueba puede ser muy útil después de construir una aplicación para identificar si su funcionamiento es el adecuado según el entorno desplegado. Si una prueba Smoke Test falla, esto significará que el software tiene un grave problema. Por tanto, no se debe desplegar cambios nuevos hasta que los fallos sean atendidos.[73]

Como se observa en la figura 34, se desarrolló la prueba de humo ejecutando el código desde la consola, donde su resultado fue la dirección donde abrirá nuestra página web. Se observa que la ampliación responde a la petición, como era de esperar.

[73] https://programacionymas.com/blog/tipos-de-testing-en-desarrollo-de-software

Figura 41. Prueba de Humo



Tomado de: Autores

Figura 42. Página Web

Tomado de: Autores

Las pruebas System Test o pruebas de sistema tienen como objetivo ejercitar pruebas que comprueben la integración de los sistemas verificando el correcto funcionamiento de las interfaces entre los distintos subsistemas que lo componen. Dentro de estas pruebas se pueden aplicar las siguientes subpruebas[74]:

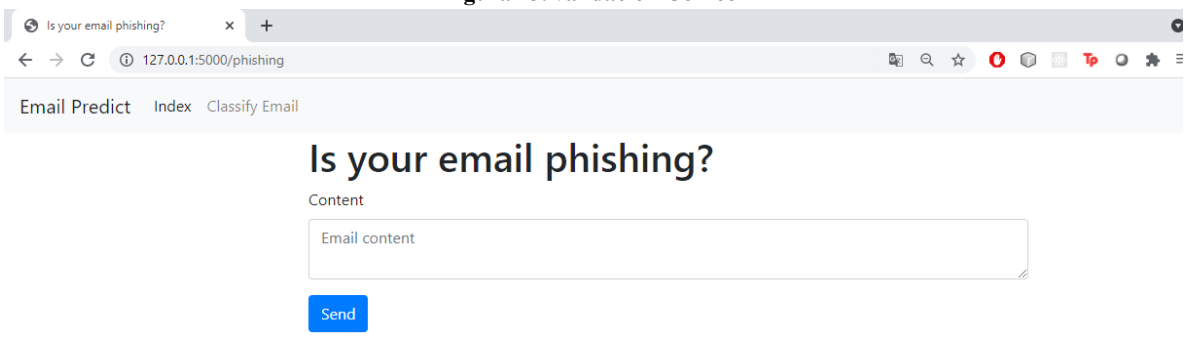
- Pruebas funcionales.
- Pruebas de comunicaciones.
- Pruebas de rendimiento
- Pruebas de volumen
- Pruebas de sobrecarga.
- Pruebas de disponibilidad de datos.
- Pruebas de facilidad de uso.

[74]Pruebas del Sistema tomado de: https://manuel.cillero.es/doc/metodologia/metrica-3/tecnicas/pruebas/sistema/#:~:text=Las%20pruebas%20del%20sistema%20tienen,con%20los%20que%20se%20comunica.

- Pruebas de operación.
- Pruebas de entorno.
- Pruebas de seguridad.

Para el desarrollo de este sprint se usaron las pruebas de facilidad de uso, que consisten en comprobar la adaptación del sistema a las necesidades del usuario, asegurando que se acomoda a un entorno habitual de trabajo, para determinar las facilidades que aporta al introducir datos en el sistema y obtener los resultados.
Como se observa en la figura 36, se ingresa a *Classify Email* para iniciar la prueba funcional

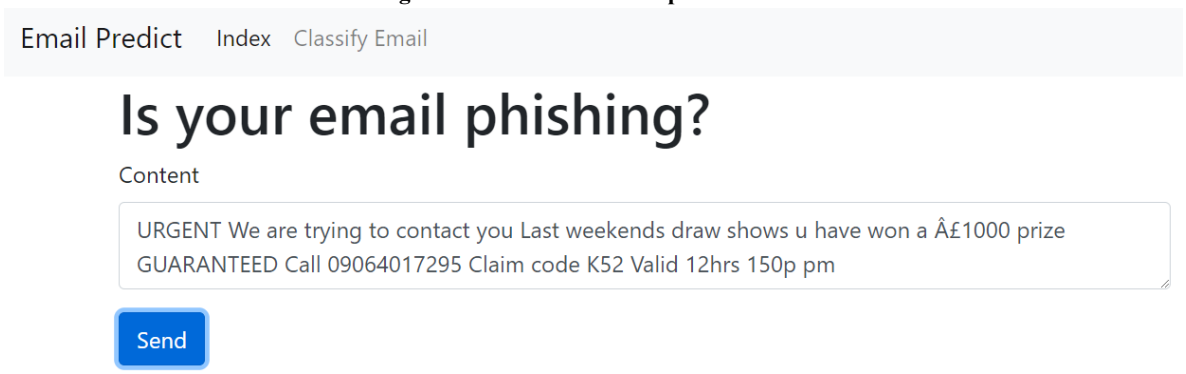
Figura 43. validación Correo



Tomado de: Autores

Luego de estar en la pantalla, se ingresa el cuerpo del correo (texto) en la sección de *Email content* en el cual se deberá dar click en send. Se observará que el texto ingresado es un correo Phishing como lo muestra en la figura 45 o en el caso de ser un correo NO Phishing se observará como en la figura 46.

Figura 44. introducir el cuerpo del correo



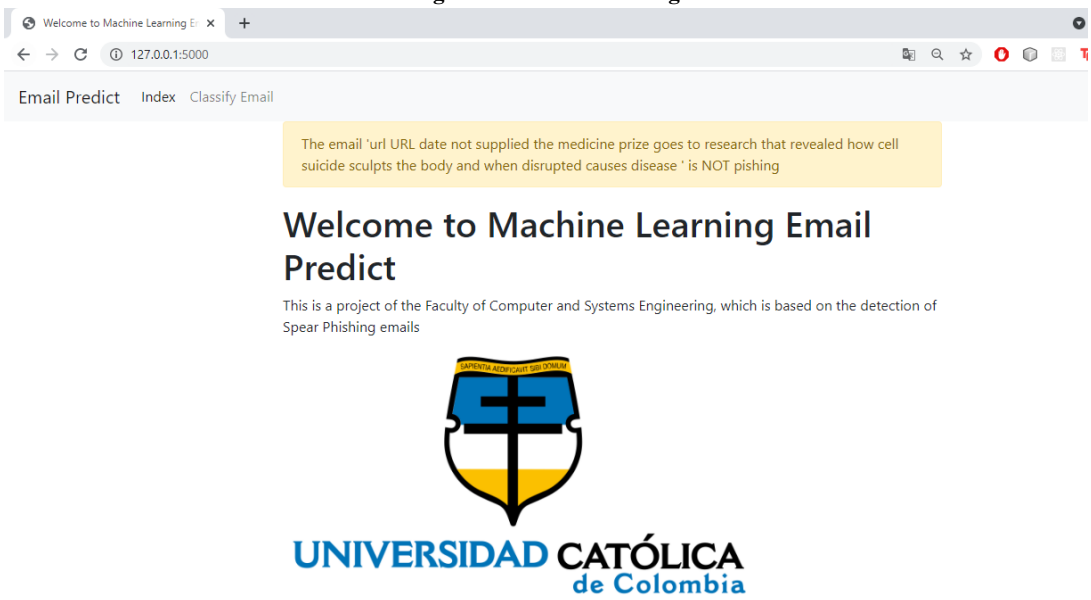
Tomado de: Autores

Figura detección correo PHISHING

Figura 45. Is Phishing

Figura detección correo no Phishing

Figura 46. Is Not Phishing



9 INSTALACIONES Y EQUIPO REQUERIDO

Las instalaciones y el equipo requeridos para el desarrollo del proyecto son los que se describen a continuación:

Un computador portátil con procesador Intel core7, memoria de 8GB de RAM y al menos 200 GB de espacio libre en el disco duro.
Anaconda con python3.6
Librerías para el experimento (pandas, numpy, operator, re, string, nltk, sklearn).
Un computador portátil con: Procesador AMD RYZEN 5 con Memoria de 12 GB de RAM.
Instalación del framework flask

10 RESULTADOS

Conforme a los objetivos estipulados, este capítulo dio a conocer un análisis de los resultados obtenidos del prototipo propuesto.

10.1 CONSTRUCCIÓN DEL CONJUNTO DE DATOS

Se elaboró un conjunto de datos a partir de correos actuales sobre estafas de phishing, adicional correos spam para obtener así un total de 3725 datos de correos Phishing y no Phishing como se muestra en la tabla 1.

Tabla 1. Conjunto de Datos

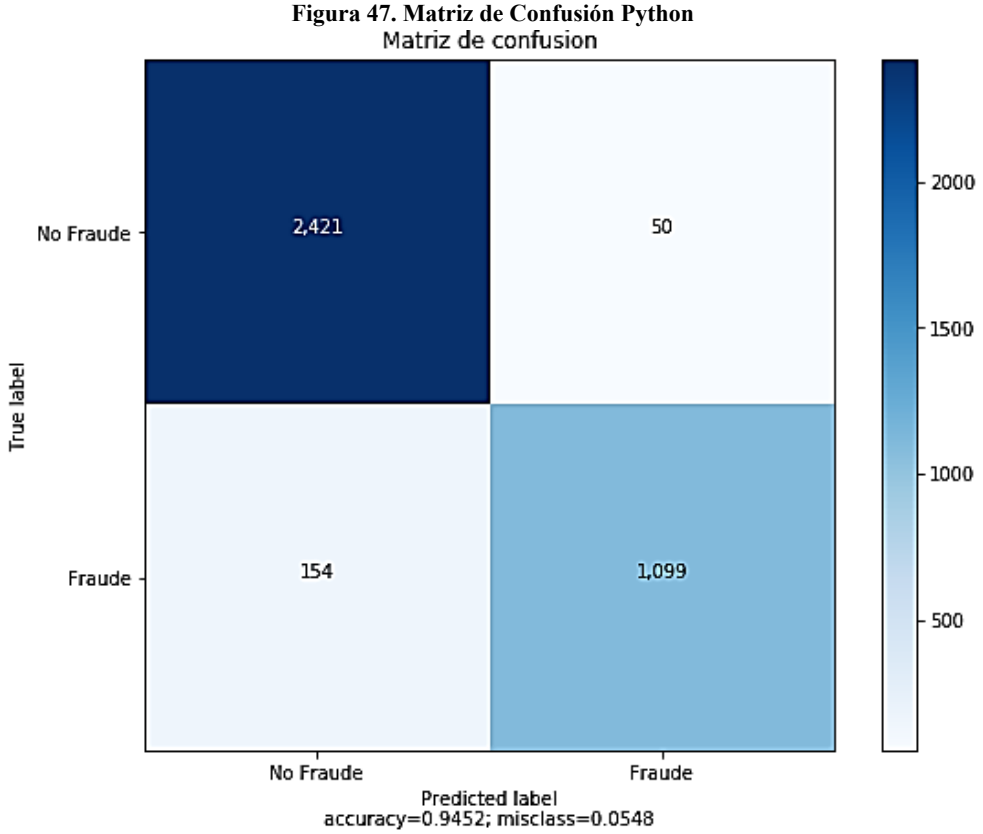
Correos	Clasificación
No Phishing (0)	2471
Phishing (1)	1254
Total	3725

Tomado de: Autores

Para utilizar el método de evaluación del modelo de clasificación se realizó una partición de los datos de entrenamiento en dos Dataset, de la siguiente manera: train 80% y test 20%.
Ahora bien, para comprobar el funcionamiento del modelo se debe hacer una medición de desempeño del algoritmo, donde identifique si los resultados obtenidos son los esperados, en este caso saber si un correo es o no phishing. Para esto se utilizó una herramienta la cual permite la medición y visualización de este desempeño: “Matriz de confusión”.

Matriz de Confusión: Permite visualizar el desempeño de un algoritmo. Cada columna de la matriz representa el número de predicciones de cada clase y las filas representan la instancia de la clase real, permite ver qué tipo de aciertos y errores está teniendo el modelo entrenado. La matriz de confusión cuenta con cuatro opciones verdadero positivo, verdadero negativo, falso negativo y falso positivo.
Al realizar en entrenamiento del modelo predictivo arboles de decisión nos generó la matriz de confusión.

Figura 47. Matriz de Confusión Python



Tomado de: Autores

En la figura 47 muestra la matriz de confusión la cual ilustra que los valores obtenidos del algoritmo demostraron lo siguiente:

Del Dataset los correos no Phishing (No fraude) que el modelo clasificó como NO Phishing fueron 2421.
De los correos No Phishing que detectó como Phishing encontró 50.
De los correos Phishing (Fraude) que no eran Phishing detectó 154
De los correos Phishing que detectó como Phishing fueron 1099.

Adicionalmente, esta matriz nos muestra la exactitud y la predicción del modelo.
La Exactitud (Accuracy)

Ecuación 8. Efectividad (Accuracy)

$$Exactitud = \frac{VP+VN}{Total} \qquad Exactitud = \frac{2421+1099}{3725}$$

$$Exactitud = 0.945$$

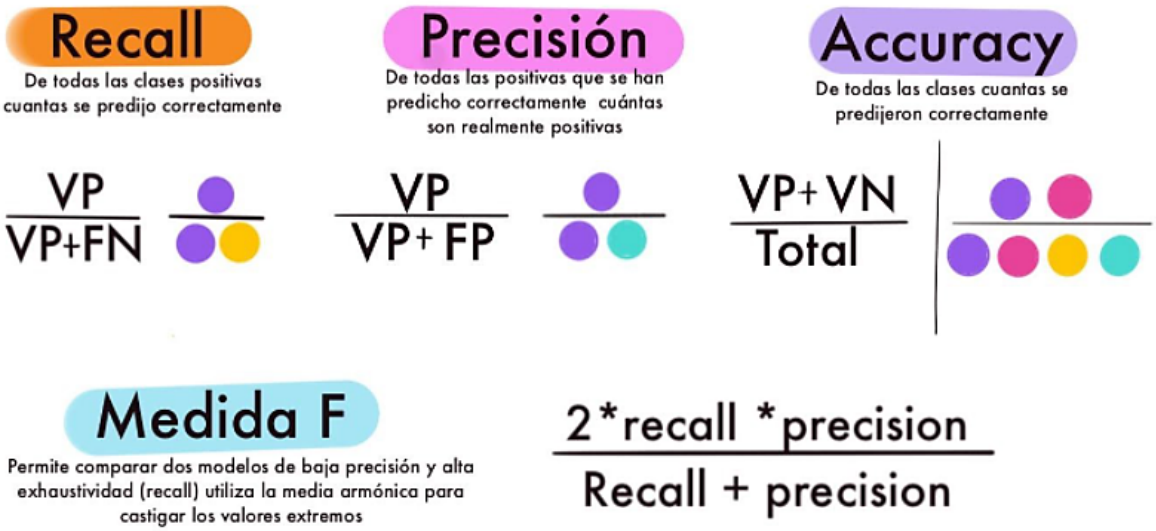
VP= Verdaderos Positivos VN= Verdaderos Negativos
FP= Falsos Positivos FN= Falsos Negativos
La tasa de error es el porcentaje del conjunto de datos que se clasificó incorrectamente. El modelo propuesto obtuvo un 0.05.

Ecuación 9. Tasa de Error Matriz de Confusión

$$Tasa\ de\ Error = \frac{FP+FN}{Total} \qquad Tase\ de\ Error = \frac{50+154}{3725}$$

$$Tasa\ de\ Error = 0.05$$

Figura 48. Explicación de Las métricas



Tomado de: https://nataliaacevedo.com/matriz-de-confusion-en-machine-learning-explicado-paso-a-paso/

En la figura 49 se muestran los valores resultantes del modelo de clasificación.

Figura 49. Precisión del Modelo

```
Precisión del modelo:
0.9456751158232217
```

Out[6]:

	precision	recall	f1-score	support
0	0.940194	0.979765	0.959572	2471.00000
1	0.956484	0.877095	0.915071	1253.00000
accuracy	0.945220	0.945220	0.945220	0.94522
macro avg	0.948339	0.928430	0.937321	3724.00000
weighted avg	0.945675	0.945220	0.944599	3724.00000

Fuente: Autores

10.2 DISEÑO DE LA ARQUITECTURA

La arquitectura de software es importante para la elaboración de un proyecto de desarrollo, debido a que es esencial tener una estructura base que comprenda la totalidad del proyecto, en otras palabras, facilita la comunicación entre todas las partes interesadas en el desarrollo. También permite tomar decisiones con mayor destreza, para así indicar el ritmo del proyecto.

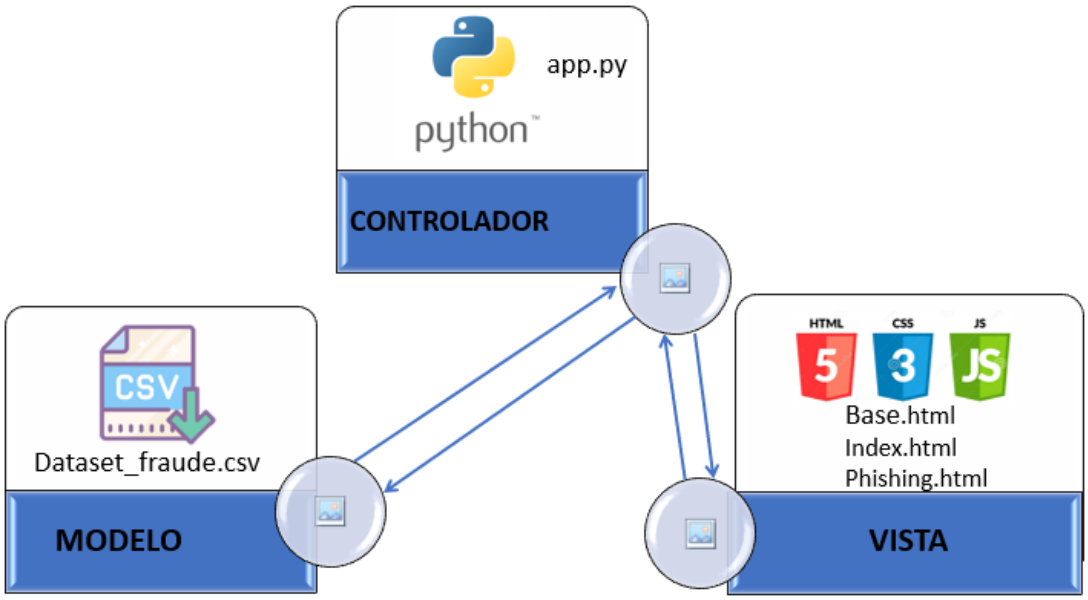
Después del realizar el análisis del desarrollo de la propuesta, en el sprint 2 de la metodología se evidenció que uno de los patrones que más se acopló al prototipo propuesto es el patrón MVC debido a que este modelo tiene ventajas como:

Separación clara entre los componentes que contiene el prototipo; esto a su vez permitirá una implementación por separado.

Este patrón tiene la ventaja de que su Interfaz de Programación es ordenada por lo tanto facilita los cambios que una persona externa quiera realizar; con esta ventaja y pensando en trabajos futuros los investigadores podrá reemplazar el Modelo, la Vista o el Controlador, sin aparente dificultad.[75]

En la figura 50 se evidencia de la estructura MVC, adicional en la figura 51 el desarrollo del prototipo propuesto.

Figura 50. MVC Desarrollo prototipo

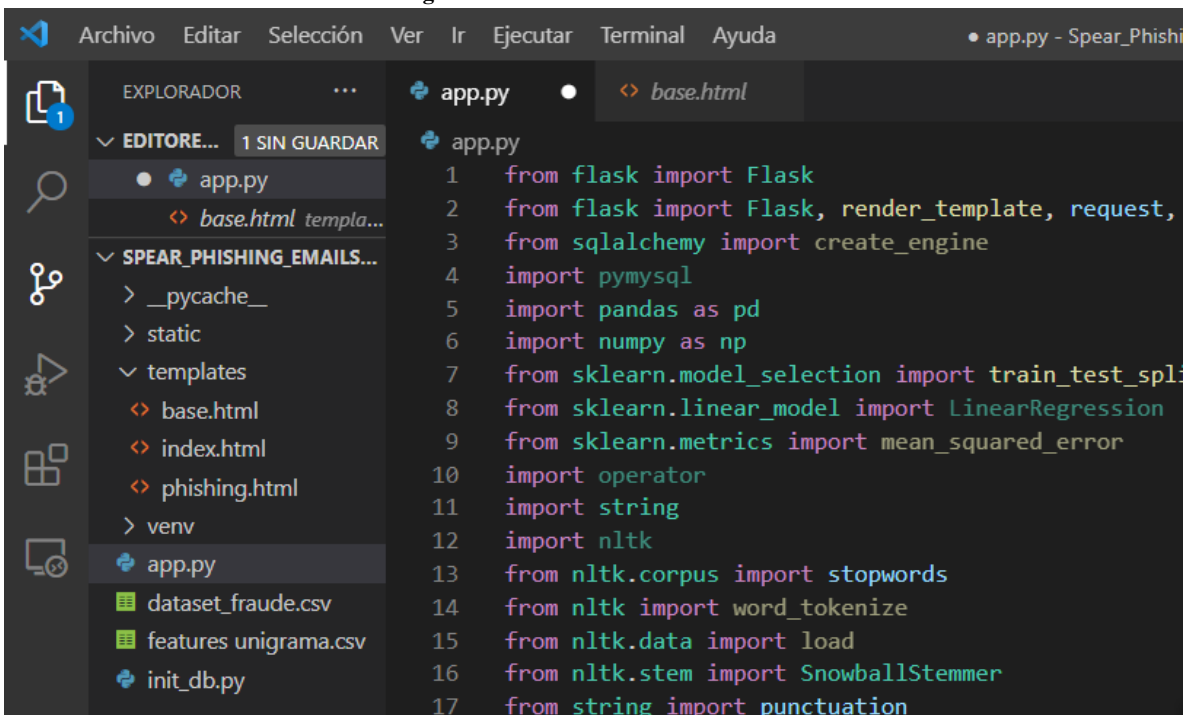


Tomado de: Autores

En el código desarrollado se observa el MVC

[75] 2012. Yenisleidy Fernández Romero, Yanette Díaz González Patrón Modelo-Vista-Controlador

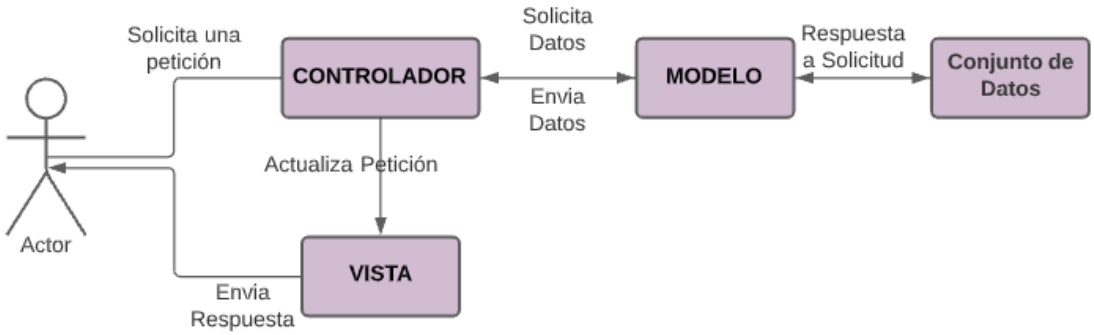
Figura 51. Vista del Desarrollo MVC



Tomado de: Autoras

En la figura 52 se muestra el diagrama de la relación que tendrá el usuario con el sistema y como éste sistema reaccionará a las peticiones.

Figura 52. Interrelación entre los elementos del patrón MCV.



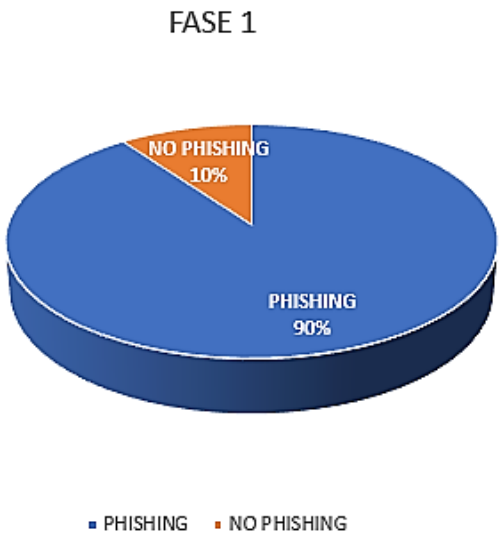
Tomado de: Autores

10.3 PRUEBAS WEB

Para este objetivo se hace uso del prototipo realizado de detección de correos phishing el cual ejecutó alrededor de 90 pruebas, las cuales se decidió dividir en 3 fases:

Fase 1: Realizar 30 pruebas con correos electrónicos catalogados como Spear Phishing

Figura 53. Correos Spear Phishing

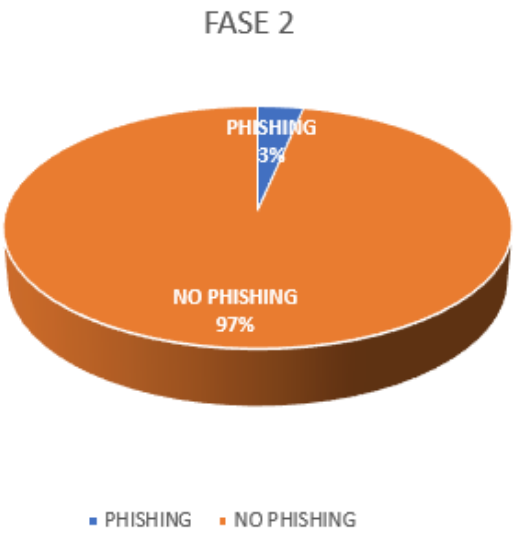


Tomado de: Autores

En la Figura 53 se evidencia que del 100% equivalente a 30 correos electrónicos Spear Phishing, fueron catalogados por el prototipo de la siguiente manera. El 90%, que equivale a 27 correos, fueron catalogados como verdaderos positivos. El restante 10%, que equivale a 3 correos, fueron clasificados como falsos negativos.

Fase 2: Realizar 30 pruebas con correos electrónicos catalogados como No Spear Phishing.

Figura 54. No Spear Phishing

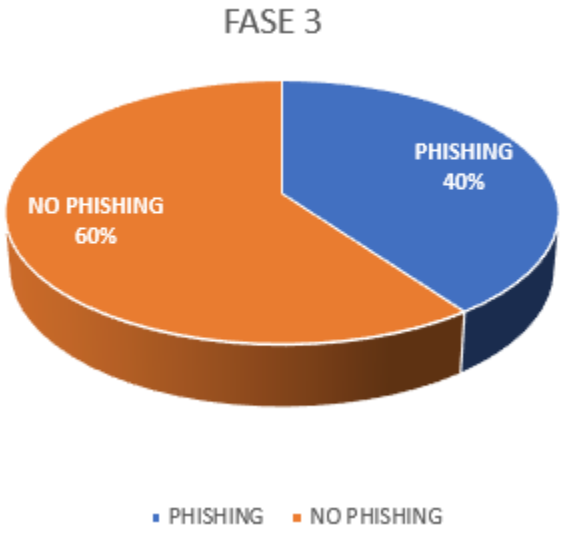


Tomado de: Autores

En la Figura 54 se evidencia que del 100% equivalente a 30 correos electrónicos no fraudulentos, fueron catalogados por el prototipo de la siguiente manera: el 97%, que equivale a 29 correos, fueron catalogados como verdaderos negativos, el restante 3%, que equivale a 1, correo fue clasificado como falso positivo.

Fase 3: En esta fase se realizó la división de correos, 15 que se catalogaron como Spear Phishing y 15 como No Spear Phishing.

Figura 55. Spear Phishing y No Spear Phishing



Tomado de: Autores

En la Figura 55 se evidencia que del 100% equivalente a 30 correos electrónicos, el 50% equivale son Spear Phishing y el otro 50% no lo son. Estos fueron catalogados por el prototipo de la siguiente manera:
El 60% que equivale a 18 correos fueron catalogados como no Phishing de los cuales 15 correos fueron clasificados como verdaderos negativos, los 3 correos electrónicos restantes los clasifico como falso negativo.
El 40% que equivale a 12 correos fueron catalogados como verdaderos positivos. Con esto se analiza que 3 correos fueron catalogados como falsos positivos.

11 CONCLUSIONES

Se concluye que este trabajo ha cumplido con el principal objetivo propuesto que es la detección de correos electrónicos Spear Phishing, aplicando métodos de categorización de texto y usando métodos de aprendizaje automático, debido a que es una de las alternativas más viables para identificar automáticamente las técnicas de ingeniería social. Por otro lado, se concluyó que el algoritmo implementado (Random forest) consiguió un resultado de precisión de un 94% de acierto en los correos electrónicos. Con esto se infiere que el algoritmo ejecutado fue el correcto.

Para el caso de detección de los correos electrónicos Spear phishing, se evidenció que es necesario contar con un Dataset reciente y con un gran número de registros, con el fin de mejorar el umbral de detección. Debido a que los ciberdelincuentes siempre buscan una técnica nueva de atrapar a las posibles víctimas.

Se evidenció que, en las pruebas realizadas, el prototipo propuesto detecta mensajes SMS, debido a que el usuario debe ingresar el cuerpo del mensaje y este algoritmo analiza las palabras más frecuentes por los ciberdelincuentes, Por este motivo el rendimiento del algoritmo se ve afectado a causa de la cantidad de palabras que debe comparar.

.

12 TRABAJOS FUTUROS

Este trabajo de grado y como en cualquier otro proyecto de investigación, existen otras líneas de investigación que están abiertas y en las que son posibles continuar con este proyecto. Durante el desarrollo de este trabajo de grado, han surgido algunas líneas futuras de trabajo, que se han dejado abiertas y que se pueden desarrollar en un futuro, algunas de ellas son:

Optimización de tiempos en el uso de la aplicación.

Expandirlo a entornos móviles.

Expandirlo a entornos de mensajes de texto.

Mejoramiento de look and feel.

Actualizar continuamente el conjunto de datos.

Aumentar el conjunto de datos.

Llevar la aplicación a casos reales.

Desplegar la página web en internet.

13 BIBLIOGRAFÍA

Abder-Rahman Ali, Introducción a pandas, {en línea}, consultado en mayo del 2021, Disponible en: https://riptutorial.com/Download/pandas-es.pdf

Ali Ghorbani, Huajie Zhang y Chair Rongxing. A PHISHING E-MAIL DETECTION APPROACH USING MACHINE LEARNING 2017

Andronicus A. Akinyelu and Aderemi O. Adewumi. Classification of Phishing Email Using Random Forest Machine Learning Technique. 2014.

AO Kaspersky Lab. Consejos para protegerse contra el cibercrimen, {en línea}, 2020, Disponible en: https://latam.kaspersky.com/resource-center/threats/what-is-cybercrime

Aprendizaje Pandas; Free unaffiliated eBook created from Stack Overflow contributors, {en línea}, Consultado Mayo del 2021, Disponible en: https://riptutorial.com/Download/pandas-es.pdf

ARCOS Sebastián Sergio, Ingeniería social Psicología aplicada a la seguridad informática, {en línea}, 1 de junio del 2011, Disponible de: https://idoc.pub/documents/ingenieria-social-1430g9713j4j

BAADEL Said y JOAN Lu, Data Analytics: Intelligent Anti-Phishing Techniques Based on Machine Learning, {en línea}, 2019, Disponible en: https://www.worldscientific.com/doi/abs/10.1142/S0219649219500059

BELCIC Ivan, Guía esencial del phishing: cómo funciona y cómo defenderse, {en línea}, 2020, Disponible en: https://www.avast.com/es-es/c-phishing

CALAMEO Keyla, Correos electrónicos, {en línea}, 2020, Disponible en https://es.calameo.com/read/004971302896b8f50921a

Ccori huaman wilber Los 10 patrones comunes de arquitectura de software {en línea} 2018 Disponible en: https://medium.com/@maniakhitoccori/los-10-patrones-comunes-de-arquitectura-de-software-d8b9047edf0b

Cillero Manuel Pruebas del sistema {en línea} Disponible en: https://manuel.cillero.es/doc/metodologia/metrica-3/tecnicas/pruebas/sistema/#:~:text=Las%20pruebas%20del%20sistema%20tienen,con%20los%20que%20se%20comunica.

Cómo reconocer y evitar las estafas de phishing de la comisión federal de comercio Información para consumidores, {en línea}, Mayo 2019, Disponible en

https://www.consumidor.ftc.gov/articulos/como-reconocer-y-evitar-las-estafas-de-phishing

Cordero vargas Jorge Universidad de Costa Rica Lematización 2009

Correo electrónico proyecto hola orientación, {en línea}, Disponible de: https://www.uv.mx/personal/rcordoba/files/2014/11/Correo_electronico.pdf

CSO Computerworld Gmail actualiza su sistema de seguridad {en línea},2017 Disponible en: https://cso.computerworld.es/tendencias/gmail-actualiza-su-sistema-de-seguridad

CSS, MDN Web Docs, {en línea}, 29 abril 2021, Disponible en: https://developer.mozilla.org/es/docs/Glossary/CSS

Explicación alternativa para accuracy, precision, recall y f1-score, {en línea}, Mayo del 2019, Disponible en: https://steemit.com/spanish/@waster/explicacion-alternativa-para-accuracy-precision-recall-y-f1-score.

Firake, S. M., Soni, P., & Meshram, B. B. Tool for Prevention and Detection of Phishing E-Mail Attacks. 2011

FLORES HERRERA Javier, ¿Qué es HTML? Código facilito, {EN LÍNEA}, 25 agosto del 2015, Disponible en: https://codigofacilito.com/articulos/que-es-html

Fraudes Bancarios, {en línea}, 2021, Disponible en https://www.larepublica.co/fraude-bancario

GALEANO Susana, Marketing Economía, {en línea}, 2020, Disponible en https://marketing4ecommerce.net/usuarios-internet-mundo/

GALLEGO YUSTE. Alberto. Delitos informáticos: Malware, fraudes y estafas a través de la red y cómo prevenirlos, {en línea}, 2012, Disponible en: https://e-archivo.uc3m.es/bitstream/handle/10016/16868/pfc_alberto_gallego_yuste.pdf?sequence=1

Gangavarapu Tushaar, Jaidhar y Bhabesh Chanduka. Applicability of Machine Learning in Spam and Phishing Email Filtering. 2020

GONZÁLEZ GARCÍA Carlos, Ciencia Cognitiva, {en línea}, 2017, Disponible en: http://www.cienciacognitiva.org/files/2017-20.pdf

GUTIERREZ Javier J, ¿Qué es framework web?, {en línea}, consultado en mayo 2021, Disponible en: http://www.lsi.us.es/~javierj/investigacion_ficheros/Framework.pdf

Hard2bit Cybersecurity, Blog - Latest News, {en línea}, 23/04/2019, Disponible en https://hard2bit.com/blog/8-tipos-de-ataque-phishing-que-ponen-en-riesgo-tu-seguridad/

Hard2bit Cybersecurity, Blog - Latest News, {en línea}, 23/04/2019, Disponible en https://hard2bit.com/blog/8-tipos-de-ataque-phishing-que-ponen-en-riesgo-tu-seguridad/

Hernandez Uriel, Codigofacilito MVC (Model, View, Controller {en línea}, Disponible en: https://codigofacilito.com/articulos/mvc-model-view-controller-explicado

Informe tendencias del cibercrimen en Colombia, Copyright Segunda Edición, {en línea}, 2020, Disponible en: https://www.ccit.org.co/wp-content/uploads/informe-tendencias-abril-2020-final.pdf

Inteligencia Artificial ¿Qué es? SALESFORCE LATINOAMÉRICA, {en línea}, 2017, Disponible en: https://www.salesforce.com/mx/blog/2017/6/Que-es-la-inteligencia-artificial.html#:~:text=La%20Inteligencia%20artificial%20es%20el,m%C3%A1quinas%20piensan%20como%20seres%20humanos%E2%80%9D.

Inteligencia artificial-Unidad 4 Redes Neuronales, {en línea}, 2020, Disponible en https://sites.google.com/site/mayinteligenciartificial/unidad-4-redes-neuronales

JUSTE Marta, Ciberataques: la amenaza aumenta, {en línea}, 2020, Disponible en:https://www.expansion.com/economía-digital/innovación/2020/05/27/5ecbaee5468aeb0f238b4599.html

La librería Pandas, {en línea}, 04 octubre 2020, Disponible en: https://aprendeconalf.es/docencia/python/manual/pandas/

LUNA Fernando, JavaScrip Aprende a programar en el lenguaje Web, {en línea}, consultado en Mayo del 2021, Disponible en: https://books.google.es/books?hl=es&lr=&id=SqikDwAAQBAJ&oi=fnd&pg=PA4&dq=que+es+html%2Bcss%2BJs&ots=pz6hW_0kFA&sig=n7WDi8D05qM4zeoCwB4GFI4bSMM#v=onepage&q=que%20es%20html%2Bcss%2BJs&f=false

MARTINEZ HERAS José, IArtificial.net 15 Librerías de Python para Machine Learning, {en línea}, 10 octubre 2020, Disponible en: https://www.iartificial.net/librerias-de-python-para-machine-learning/

MARTINEZ, El cibercrimen no descansa, estas son las proyecciones para el 2020, El tiempo, {en línea}, 2019, Disponible en: https://www.eltiempo.com/tecnosfera/dispositivos/cifras-de-ciberataques-de-2019-y-tendencias-para-el-2020-435508

MEDINA BRANCH Rosgaby, Estadísticas de la situación digital de Colombia en el 2019 y 2020, {en línea}, 27 de Abril del 2020, Disponible en: https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-colombia-en-el-2019-y-2020/#:~:text=En%20un%20a%C3%B1o%20(del%202019,crearon%203.4%20millones%20nuevos%20perfiles

Méndez José, Riverola Florentino, Díaz Fernando,Corchado Juan Sistemas inteligentes para la detección y filtrado de correo spam: una revisión 2007

Métodos de petición HTTP, MDN Web Docs, {en línea}, 14 mayo 2021 Disponible en: https://developer.mozilla.org/es/docs/Web/HTTP/Methods

National Cyber security Centre, {en línea}, 15 de octubre del 2015, Disponible en: https://www.ncsc.gov.uk/information/how-cyber-attacks-work

Nish Anil Implementación de comunicación basada en eventos entre microservicios {en línea} 2021, Disponible en: https://docs.microsoft.com/es-es/dotnet/architecture/microservices/multi-container-microservice-net-applications/integration-event-based-microservice-communications.

Ona Diego, Zapata Lenín, Fuertes Walter, Rodríguez German, Eduardo Benavides y Theofilos Toulkeridis. Phishing Attacks: Detecting and Preventing Infected E-mails Using Machine Learning Methods. 2019.

PAGNOTTA Sabrina, Las víctimas de ciberataques perdieron 1,33 mil millones de dólares en 2016, {en línea}, 2017, Disponible en https://www.welivesecurity.com/la-es/2017/06/28/victimas-ciberataques-millones-dolares/

PASCUAL ESTAPÉ Juan Antonio, “Inteligencia artificial-Computer Hoy”, {en línea}, 24 de agosto del 2019, Disponible en https://computerhoy.com/reportajes/tecnologia/inteligencia-artificial-469917

Programacionymas {en línea} Disponible en: https://programacionymas.com/blog/tipos-de-testing-en-desarrollo-de-software

¿Qué son archivos adjuntos de correo?, {en línea}, 2020, Disponible en https://www.ionos.es/ayuda/correo/glosario-explicaciones-sobre-conceptos-y-temas-importantes/archivos-adjuntos-de-correo/

RUSSO Claudia, RAMÓN Hugo, ALONSO Nicolás, CICERCHIA Benjamín, ESNAOLA Leonardo, TESSORE Juan Pablo, Tratamiento masivo e Datos Utilizando técnicas e Machine Learning,{en línea}, Consultado el 3 de abril del 2021,Disponible en: https://digital.cic.gba.gob.ar/bitstream/handle/11746/5603/11746_5603.pdf-PDFA.pdf?sequence=1&isAllowed=y

SAMPEDRO Javier, El desarrollo de la inteligencia artificial da vértigo en los círculos económicos. ¿Van los robots a dejarnos sin trabajo?, {en línea}, 8 de junio del 2016, Disponible en https://elpais.com/elpais/2016/06/08/opinion/1465383749_599768.html

Scikit-learn Machine Learning in Python, {en línea}, Consultado en abril 2021, Disponible en: https://scikit-learn.org/stable/

Secretaria general del estado, Constitución Política Colombiana, {en línea}, 26 de abril del 2021, Disponible en: http://www.secretariasenado.gov.co/index.php/constitucion-politica

SEGUÍ PAREJO Eric, Publisuites, {en línea}, consultado en Mayo del 2021, Disponible en: https://www.publisuites.com/blog/tf-idf/#Que_es_el_TF_IDF

Segura Ariel Alejandro, Arquitectura de Software de Referencia para Objetos Inteligentes en Internet de las Cosas

Significado de URL,Tecnología e Innovación, {en línea}, consultado el 20 de marzo del 2021, Disponible en https://www.significados.com/url/#:~:text=URL%20son%20las%20siglas%20en,significa%20Localizador%20Uniforme%20de%20Recursos.&text=As%C3%AD%2C%20hay%20un%20URL%20para,por%20primera%20vez%20en%201991

¿Somos conscientes de los retos y principales aplicaciones de la Inteligencia Artificial?, Iberdrola, {en línea}, 2020, Disponible en: https://www.iberdrola.com/innovacion/que-es-inteligencia-artificial.

Software del sol popups o ventanas emergentes, {en línea}, 2021, Disponible en https://www.sdelsol.com/glosario/popups-o-ventanas-emergentes/

Srishti Rawal,Bhuvan Rawal,Aakhila Shaheen y Shubham Malik. Phishing Detection in E-mails using Machine Learning.2017

State of the Phishing, an in-depth look at user awareness, vulnerability and resilience, { en línea}, 2020, Disponible en: https://www.proofpoint.com/sites/default/files/gtd-pfpt-uk-tr-state-of-the-phish-2020-a4_final.pdf

Suraj J Pai,Rakshitha Gokuldas, Rahul Kakkadan, Sourabh Hegde, Ms. Saritha Suvarna. PHISHING WEBSITE ANALYZER TO SECURE E-BANKING AND E-COMMERCE WEBSITES2020
Telem@tica Fernández romero Yenisleidy, Díaz González Yanette Patrón Modelo-Vista-Controlador En: Vol. 11. No. 1, enero-abril, 2012

What is an Unsolicited Bulk Email? – Basic, {en línea}, 19.03.2021, Disponible en: https://sendpulse.com/support/glossary/unsolicited-bulk-email

14 ANEXOS

Anexo A: Conjunto de Datos

Se realiza la entrega del conjunto de datos construido para el desarrollo del proyecto en formato CSV.

Anexo B: Repositorio del Código

En el anexo B se entrega el repositorio subido a Github, con el desarrollo completo del prototipo de detección de correos electrónicos Spear Phishing.

https://github.com/SpearPhishingEmail/Spear_Phishing_emails

Anexo C: Manual de Usuario

Se realiza entrega del Manual de usuario para su correcto uso.

UNIVERSIDAD CATÓLICA
de Colombia

IMPLEMENTACIÓN DE UN PROTOTIPO FUNCIONAL DE APRENDIZAJE DE MAQUINA PARA IDENTIFICAR CORREOS ELECTRÓNICOS DE SPEAR PHISHING

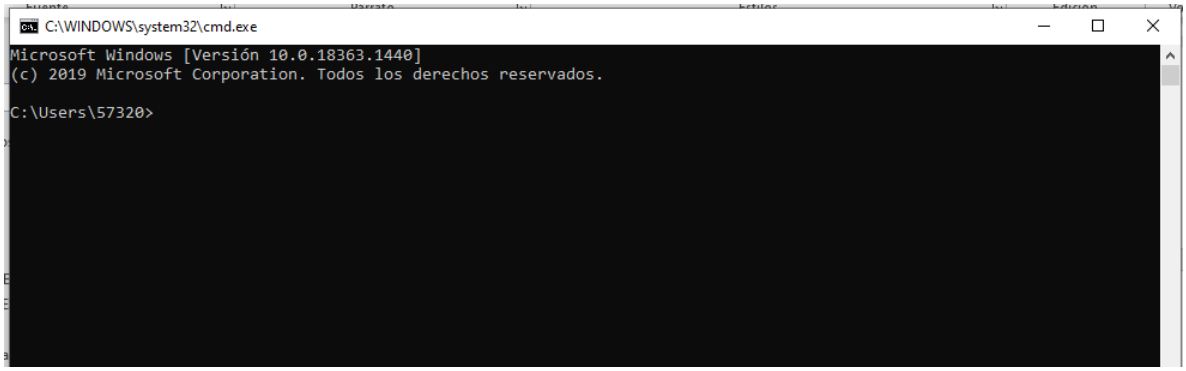
ANEXO C
MANUAL DE USUARIO

MARIA ALEJANDRA SUAREZ SANCHEZ 67000205
JHINDY HASLEYDE PARDO RODRIGUEZ 67000022

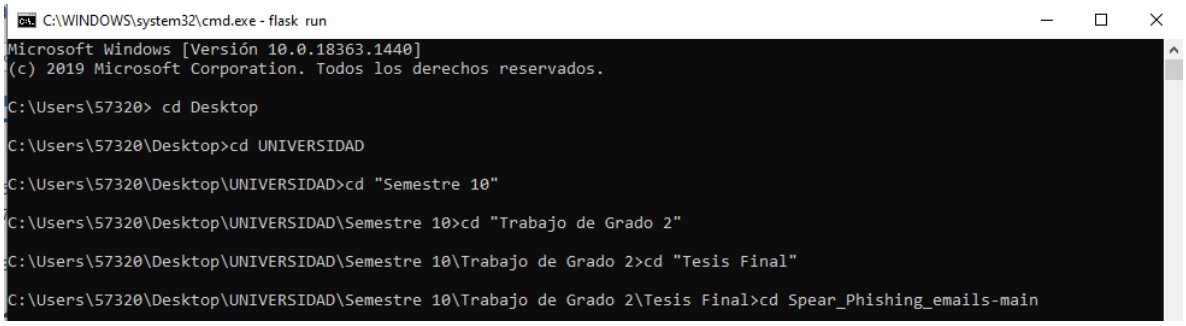
PROGRAMA DE INGENIERÍA DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD CATÓLICA DE COLOMBIA
BOGOTÁ, MAYO
2021

En este Manual se pretende explicar el funcionamiento de este prototipo web que detecta el contenido de correos electrónicos, validando si es Spear Phishing.

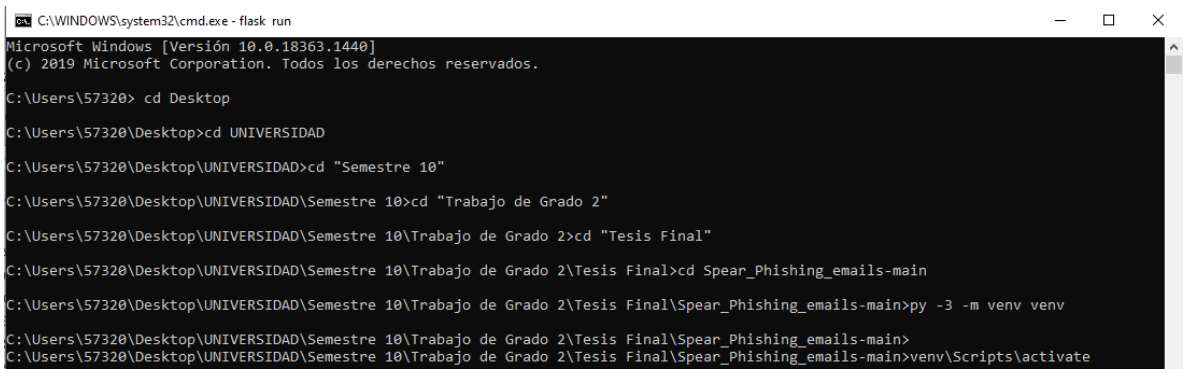
Para Ingresar al sistema el usuario deberá entrar a la consola Windows (CMD)



Una vez el usuario este en la consola deberá ubicar el documento.

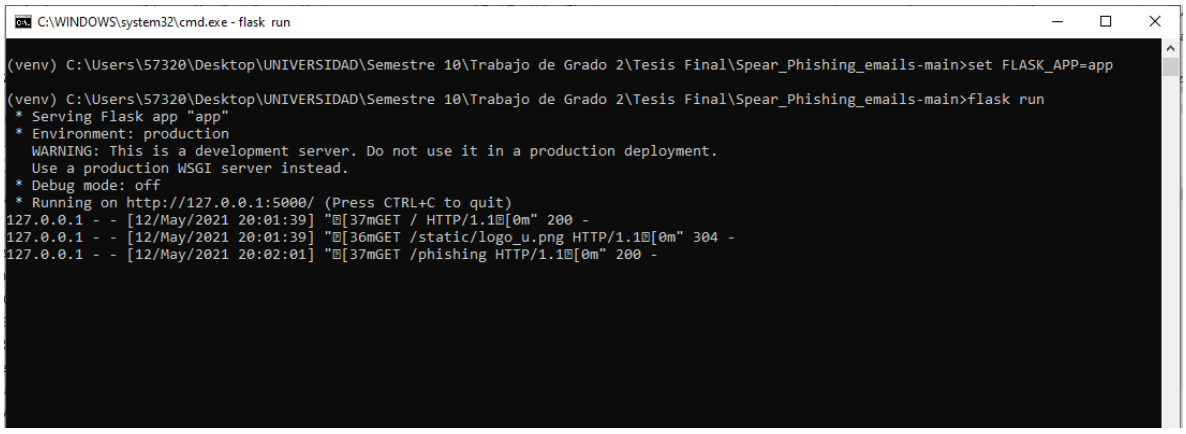


Ingresar a la carpeta donde guardó el proyecto. En este se deberá digitar en la consola *py -3 -m venv venv* este venv módulo es para crear un entorno virtual.
Una vez instalado el modulo para el entorno virtual se debe activar el proyecto, para esto se deberá usar *venv\Scripts\activate*
Ingresar al siguiente enlace donde encontrará los comandos:
https://flask.palletsprojects.com/en/1.1.x/cli/



Para poder correr el flask y así observar la página deberá ingresar

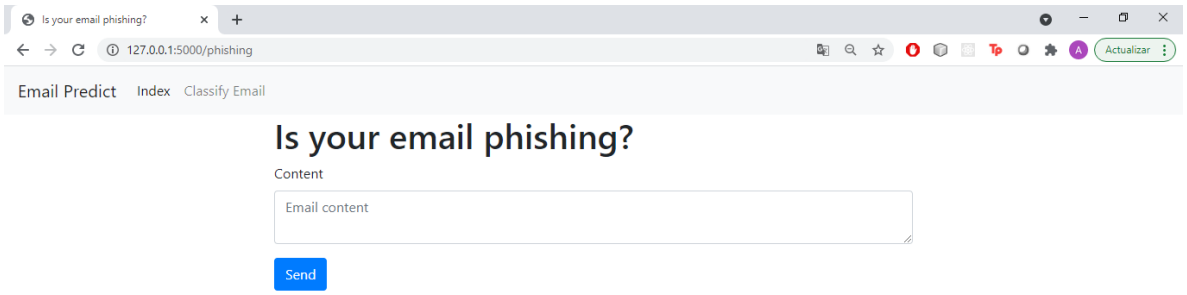
set FLASK_APP=app y *flask run*
En el siguiente link encontrará los comandos:
https://flask.palletsprojects.com/en/1.1.x/cli/

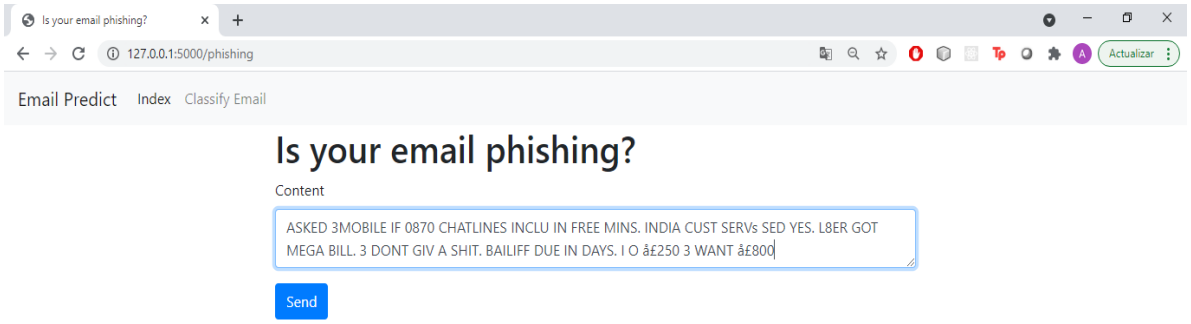


Una vez carga la consola CMD, esta mostrará la URL que deberá ingresar en el buscador Google.
Una vez ingresado la URL observará el escudo de la Universidad Católica de Colombia y una descripción acerca de Phishing.

En esta página encontrará en la parte superior una pestaña *Classify Email*. En esta pestaña el usuario deberá dar click.

Esta ventana contiene un cuadro de texto. En este el usuario deberá ubicar el cuerpo del correo que desea revisar. Para iniciar la detección de si el correo es Spear Phishing el usuario dará click en la parte inferior en el botón *Send*





Nota: El programa tiene un tiempo de respuesta según el tamaño del correo.

El usuario observará la respuesta de la detección en el recuadro naranja este dirá si es Phishing finalizando el cuerpo del correo como se muestra en la imagen