Algoritmo Machine Learnig en los sistemas de filtrado caso práctico SPAM de google en las cuentas de correo Institucionales de la F.A.F.I.

Algoritmo Machine Learning of the filtering systems of google SPAM practical cases in the Institutional mail accounts of the F.A.F.I.

Ernesto David Mora Suarez[1,*]
Universidad Técnica de Babahoyo
{davidernestoms@gmail.com}
Fecha de recepción: 7 de Julio de 2018 - **Fecha de revisión:** 25 de Julio de 2018

Resumen

El presente artículo analiza el funcionamiento del algoritmo Machine Learning en los sistemas de filtrados de correos SPAM adoptado por la compañía Google Corp. en las cuentas institucionales de Universidad Técnica de Babahoyo, tomando como campo de acción la Facultad de Administración Finanzas e Informática.

Para efectuar la investigación, se ha adoptado de las metodologías descriptiva y experimental de laboratorio, para evidenciar los efectos que presentan los correos no deseados o también conocidos como SPAM. Se empleó encuestas a los usuarios que poseen cuenta institucional, así como también cuentas personales. Se efectuó pruebas de laboratorio, para demostrar el efecto de las cuentas google ante ataques SPAM.

Palabras claves **– SPAM, algoritmo, google.**

Summary

This article analyzes the functioning of the Machine Learning algorithm in the SPAM email filtering systems adopted by the company Google Corp. in the institutional accounts of the Technical University of Babahoyo, taking the Faculty of Administration, Finance and Information as a field of action.

To carry out the research, descriptive and experimental laboratory methodologies have been adopted to demonstrate the effects of unwanted emails or also known as SPAM. Surveys were used for users who have an institutional account, as well as personal accounts. Laboratory tests were carried out to demonstrate the effect of google accounts against SPAM attacks.

Keywords **– SPAM, algorithm, google.**

INTRODUCCIÓN

La mayoría de internautas gozan de las ventajas de los servicios web sin verse afectados su información, equipos y comunicaciones, gracias a aplicaciones que minimizan el riesgo tecnológico no autorizado, siendo el caso de la protección de correos por filtros SPAM.

El presente artículo analiza el funcionamiento del algoritmo **Machine Learning** en los sistemas de filtrados de correos SPAM adoptado por la compañía Google Corp. en las cuentas institucionales de Universidad Técnica de Babahoyo, tomando como campo de acción la Facultad de Administración Finanzas e Informática.

Para efectuar la investigación, se ha adoptado de las metodologías descriptiva y experimental de laboratorio, para evidenciar los efectos que presentan los correos no deseados o también conocidos como SPAM. Se empleó encuestas a los usuarios que poseen cuenta institucional, así como también cuentas personales. Se efectuó pruebas de laboratorio, para demostrar el efecto de las cuentas google ante ataques SPAM.

Finalmente se concluye que uno de los mejores agentes y servidor de correos empleados a nivel mundial, por su efectividad y goce casi inmune es atribuida a la firma Google Corp; a través, de su aplicativo open source conocido como TensorFlow que trabaja con Machine Learning.

DESARROLLO

En los años 1950 Alan Turing crea el "Test de Turing" para determinar si una máquina era realmente inteligente. Para pasar el test, una máquina tenía que

*Estudiante de la Universidad Técnica de Babahoyo

ser capaz de engañar a un humano haciéndole creer que era humana en lugar de un computador y unos años más tarde Arthur Samuel escribe el primer programa de ordenador capaz de aprender. (Pacheco, 2017)

El software era un programa que jugaba a las damas y que mejoraba su juego partida tras partida, mientras que Martin Minsky y John McCarthy, con la ayuda de Claude Shannon y Nathan Rochester, organizan la conferencia de Darthmouth de 1956, considerada como el evento donde nace el campo de la Inteligencia Artificial. Durante la conferencia, Minsky convence a los asistentes para acuñar el término "Artificial Intelligence" como nombre del nuevo campo. (Pacheco, 2017)

Los años 80 estuvieron marcados por el nacimiento de los sistemas expertos, basados en reglas. Estos fueron rápidamente adoptados en el sector corporativo, lo que generó un nuevo interés en Machine Learning, ya entonces a finales de los 80, y durante la primera mitad de los 90, llegó el segundo "Invierno" de la Inteligencia Artificial. Esta vez sus efectos se extendieron durante muchos años y la reputación del campo no se recuperó del todo hasta entrado los 2000. (Pacheco, 2017)

A fecha de hoy estamos viviendo una tercera explosión de la inteligencia artificial. Aunque existen escépticos que no descartan un posible tercer invierno, esta vez los avances del sector están encontrando aplicabilidad en empresas hasta el punto de crear mercados enteros y producir cambios significativos en la estrategia de grandes y pequeñas empresas. (Pacheco, 2017)

La gran disponibilidad de datos parece ser el fuel que está alimentando los motores de los algoritmos que, a su vez, han roto las limitaciones de cálculo que existían antes de la computación distribuida. Todo parece indicar que seguiremos disponiendo de más y más datos con los que alimentar nuestros algoritmos mientras que la comunidad científica no parece quedarse sin ideas con las que seguir avanzando en el campo. 2017 y los próximos años prometen ser realmente frenéticos. (Pacheco, 2017)

Analizar la función que cumple el algoritmo Machine Learning en los Sistemas de Filtrado de Correo Spam de Google y Demostrar cómo incide en las cuentas de correo institucional ante ataques SPAM dentro de la Universidad Técnica de Babahoyo.

Unidad de análisis

El Machine Learning en su uso más básico es la práctica de usar algoritmos para parear datos, aprender de ellos y luego ser capaces de hacer una predicción o sugerencia sobre algo. (RODRIGUEZ, 2017).

La seguridad en internet está en entredicho tras los últimos ataques y descubrimientos de malware, pero servicios como Gmail se empeñan en mantenerse a la vanguardia para también mantenerse por delante de los hackers. Y Google, experta en inteligencia artificial y aprendizaje de máquina, se ha decidido a que todo software posea sus propias herramientas de toma de decisiones, por eso introducirá su inteligencia artificial en todos los servicios, incluido Gmail. Ahora estaremos protegidos del phishing gracias al "machine learning". (Linares, 2017)

Qué es Google SafeBrowsing

Google diariamente hace un barrido de nuevos sitios web que aparecen en la red de internet. Algunos de esos sitios web pueden tener algún software malicioso que pueda ocasionar daños al usuario que navega en el citado sitio. Google SafeBrowsing es un servicio de Google para evitar contenidos no seguros. (Fernández, s.f.)

Estos sitios no seguros se pueden dividir en dos categorías:

Sitios de software malicioso que se instala en los navegadores de los usuarios para capturar información privada y confidencial.

Sitios de suplantación de identidad, que fingiendo ser lo que no son se hacen con nombres de usuario y contraseña. (Fernández, s.f.)

Google, en ese rastreo, cuando advierte alguno de estos riesgos lo pone en conocimiento de los Web master de la página web, a fin de que éstos proporcionen soluciones a los problemas observados y constatados. Obviamente, la penalización sino hay una respuesta inmediata a dichos problemas es la penalización de google en la indexación de ese sitio web. (Fernández, s.f.)

Según (Cormack, 2011) Desarrollaron un método personalizado de priorización de correo electrónico (PEP) que se enfoca especialmente en el análisis de redes sociales personales para capturar grupos de usuarios y obtener características enriquecidas que representan los roles sociales desde el punto de vista del usuario particular , así como también desarrollaron un marco de clasificación supervisado para modelar las prioridades personales sobre los

mensajes de correo electrónico, y para predecir niveles de importancia para los mensajes nuevos. (Guzella, 2009)

Propuso un modelo inspirado en el sistema inmune, llamado sistema inmune innato y adaptativo artificial (IA-AIS) y aplicado al problema de la identificación de mensajes de correo electrónico no solicitado (SPAM). Integra entidades análogas a macrófagos, y linfocitos, que modelan los sistemas inmune innato y adaptativo. Una implementación del algoritmo fue capaz de identificar más del 99% de los mensajes legítimos o SPAM en configuraciones de parámetros particulares. Se comparó con una versión optimizada del ingenuo Bayesclassifier, que obtuvo tasas de clasificación extremadamente altas. Se ha concluido que IA-AIS tiene una mayor capacidad para identificar mensajes SPAM, aunque la identificación de mensajes legítimos no es tan alta como la del ingenuo Bayesclassifier implementado.

Método clasificador de Naïve BayesEn 1998 se propuso el clasificador Naïve Bayes para el reconocimiento de spam. El clasificador bayesiano está trabajando en los eventos dependientes y la probabilidad de que ocurra un evento en el futuro que pueda detectarse a partir de la ocurrencia previa del mismo evento [12]. Esta técnica se puede utilizar para clasificar los correos electrónicos no deseados; las probabilidades de palabras juegan la regla principal aquí. Si algunas palabras ocurren a menudo, pero no en jamón, este correo electrónico entrante probablemente sea correo no deseado. Naïve bayes classifiertechnique se ha convertido en un método muy popular en el software de filtrado de correo. El filtro Bayesiano debe estar entrenado para funcionar de manera efectiva. Cada palabra tiene cierta probabilidad de aparecer en correo electrónico spam o ham en su base de datos. Si el total de las probabilidades de palabras excede un cierto límite, el filtro marcará el correo electrónico en cualquiera de las categorías. Aquí, solo se necesitan dos categorías: spam o jamón. Casi todos los filtros de spam basados en estadísticas usan el cálculo de probabilidad bayesiano para combinar las estadísticas de singletoken con un puntaje general [1], y toman una decisión de filtrado basada en el puntaje. La estadística que más nos interesa para un token T es su correo no deseado (spam rating) [10], que se calcula de la siguiente manera: Donde CSpam (T) y CHam (T) son el número de mensajes spam o de token T, respectivamente. Para calcular la posibilidad de un mensaje M con tokens {T1, ..., TN}, es necesario combinar el correo no deseado del token individual para evaluar el mensaje basura en general. La forma más sencilla de hacer clasificaciones es calcular el producto de los mensajes basura individuales y compararlo con el producto del token individual. El mensaje se considera correo no deseado si el producto de correo no deseado global S [M] es más grande que el producto H [M]. La descripción anterior se usa en el siguiente algoritmo [10]: Etapa1. TrainingParse cada correo electrónico en sus tokens constituyentes Generar una probabilidad para cada token WS [W] = Cspam (W) / (Cham (W) + Cspam (W)) almacenan los valores de spam en una base de datosStage2. FiltradoPara cada mensaje Mwhile (M no final) doscan mensaje para el próximo token Tiquery la base de datos para correo no deseado S (Ti) calcula probabilidades de mensaje acumuladasS [M] y H [M] Calcule la indicación general de filtrado de mensajes por: I [M] = f (S [M], H [M]) f es una función dependiente del filtro, como S [T] = C Spam (T) C Spam (T) + C Jamón (T) I [M] = 1 + S [M] -H [M] 2I (H [M] =? (1- S [T])) I = 1N(Libro MachineLearningSpam)

El aprendizaje detrás de Gmail Priority Inbox

Modelos

Utilizamos modelos de regresión lineal simple para seguir aprendiendo y predicción escalable. Existe una abundancia de datos para el aprendizaje de un modelo global, pero la escasez de datos existe para el aprendizaje de modelos de usuario personalizada. Utilizamos una forma sencilla de transferir el aprendizaje, donde la predicción final es la suma del modelo global y el modelo de usuario log odds (Fig. 1):

$$s = \sum_{i=1}^{n} f_i g_i + \sum_{i=|1}^{n+k} f_i w_i, \qquad p = \frac{1}{1+\exp^{-s}}.$$

(Yang., 2010)

El número de características que es denotado por n. Nosotros usamos K para características específicas de usuario adicionales que no están presentes en el modelo global. El modelo global pesos son g y se actualizan de forma independiente y mantiene fija durante las actualizaciones de modelos personales. Así, el modelo pondera personales w sólo representan cómo el usuario es diferente del modelo global de importancia. Esto se traduce en más modelos de usuario compacto y la capacidad de adoptar rápidamente los cambios en el modelo global, por

ejemplo, cuando se agregan nuevas características.

Realizamos actualizaciones en línea agresiva pasiva con el PA-II variante de regresión para combatir el alto grado de ruido en el conjunto de entrenamiento. Cada correo se usa una vez para actualizar el modelo global, y una vez para actualizar el modelo para el destinatario del mensaje, por ejemplo, la actualización de su modelo de usuario es:

$$w_i \leftarrow w_i + f_i \frac{\operatorname{sgn}(e)\max(|e| - \epsilon, 0)}{\|\mathbf{f}\|^2 + \frac{1}{2C}},$$

Desde la cual es el error C es un parámetro de regularización que sintoniza la "agresividad" de las actualizaciones y la tolerancia a la pérdida de la bisagra, o el grado de "pasividad". En la práctica, nos abuse C ajustando por correo electrónico a representar nuestra confianza en la etiqueta, por ejemplo, una corrección manual por parte de un usuario se le concede un valor superior of C. Modelos de usuario también tienen higher C que el modelo global, y los nuevos modelos de usuarios todavía tienen los valores más altos para promover el aprendizaje inicial. (Koby Crammer, 2010)

El SPAM o Correo basura

El SPAM es un problema que cada día se hace más común. Estos correos electrónicos no solicitados ni esperados llegan constantemente y aumentan el tráfico en la red Internet del Recinto. Generalmente estos emails contienen promociones, fotos, información falsa y archivos enormes que en la mayoría de los casos son virus que afectan al sistema o su computadora personal. El inglés representa ahora el idioma del 80% del spam que recibimos, y otros en el idioma español se hacen cada vez más comunes. No hay por qué alarmarse al recibirlos pues la mayoría son promociones de mercadeos de compañías que desean vender o mercadear algún producto. (RICO, 2012)

Guías para controlar el SPAM

- ¡Bórrelos!
- Verifique su correo electrónico a diario para evitar la acumulación y evitar la llegada de nuevos emails.
- Aprenda a utilizar algún programa que clasifique estos mensajes y los elimine automáticamente.
- Evite el reenvió de estos mensajes tales como donaciones, alerta de virus, amistad, amor, religiosos o de negocio.
- Cuide a quien da su email, utilice el servicio de correo electrónico para asuntos académicos, administrativos y aprenda a usar el servicio de su preferencia correctamente.
- Evite el uso de "Outlook express" sin criterios o algún tipo de protección, el SPAM puede adquirir todos sus contactos y reenviar el mensaje automáticamente, como si fuese usted mismo. (RICO, 2012).

MÉTODOS E INSTRUMENTOS DE RECOLECCIÓN DE INFORMACIÓN

Metodología Experimental

Algoritmo Machine Learnig en los sistemas de filtrado caso práctico SPAM de google en las cuentas de correo de la FAFI

La investigación se centra en la metodología Experimental. Para el investigador es una tarea crucial, y un cometido ineludible "interpretar" los experimentos con la finalidad de determinar cuál es la frecuencia de SPAM de google en los correos institucionales de la Facultad de Administración, Finanzas e Informática (Estebaranz, 2013).

Este método Según define la observación como la técnica de investigación básica, sobre el que sustenta todas las demás ya que establece la relación básica entre el sujeto que observa y el objeto que es observado, que es el inicio de toda comprensión de la realidad.

Actualmente cada estudiante posee una cuenta institucional de Google en la FAFI y como consecuencia al gran número de estudiantes, los correos SPAM pueden ocasionar problemas de virus o perdida de información personal ya que puede ser un virus encriptado o personas que quieren conocer los datos de estudiantes.

Utilizamos este método porque mediante la experimentación y observación nos permite obtener el mayor número de datos y la información que se necesitada para darle una solución a nuestro proyecto, mediante el método de experimentación se pude elaborar varias hipótesis que nos ayudara a esclarecer la situación actual de nuestro problema. (Acero, 2017).

Por ello, nos importa el estudio, como se observa habitualmente, la investigación implica la hipótesis precisa.

ANÁLISIS DE LA INFORMACIÓN

Población y Muestra

La siguiente población estuvo conformada por 2406 estudiantes, 104 docentes y 17 personal

administrativo de la Facultad de Administración, Finanzas e Informática.

Cálculo del tamaño de la muestra conociendo el tamaño de la población

La fórmula para calcular el tamaño de muestra cuando se conoce el tamaño de la población es la siguiente:

En donde,

N= Tamaño de la población

n= Muestra

E= Error máximo admisible en términos de proporción.

1= Constante.

$$n = \frac{N}{E2\,(N-1)+1}$$

Cuadro 1: Población y Muestra

ESTRATOS	POBLACION	MUESTRA
Estudiantes	2406	410
Docentes	104	86
Personal Administrativo	17	17

Elaborado por: Estudiantes de Octavo Sistemas- Vespertino

De donde la muestra de estudiantes con las que se realiza la presente investigación es de 410.

Para obtener la muestra de los docentes, seguimos el mismo procedimiento y tenemos: Entonces con los docentes, la muestra seleccionada es de 86.

Del personal administrativo se seleccionó toda la población debido a que es una cantidad pequeña la muestra es de 17.

ANÁLISIS E INTERPRETACIÓN DE DATOS

Encuesta dirigida a los estudiantes de la Facultad de Administración Finanzas e Informática.

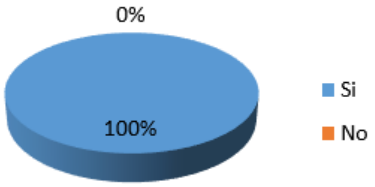
Pregunta # 1 ¿Posee una cuenta de correo institucional o en Gmail?

Tabla Nº 1

ALTERNATIVA	FRECUENCIA	PORCENTAJE
Si	410	100 %
No	0	0%
TOTAL	**410**	**100 %**

Fuente: Facultad de Administración, Finanzas e Informática.
Realizado Por: Ernesto Mora

Gráfico Nº 1



Interpretación

De 410 estudiantes encuestados se obtuvo que el 100% de ellos posee una cuenta de correo institucional o en Gmail.

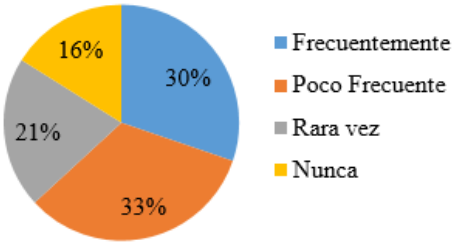
Pregunta # 2 ¿Con qué frecuencia usted lee sus correos y/o actualiza su bandeja de entrada?

Tabla Nº 2

ALTERNATIVA	FRECUENCIA	PORCENTAJE
Frecuentemente	124	30 %
Poco Frecuente	135	33 %
Rara vez	85	21 %
Nunca	66	16 %
TOTAL	**410**	**100 %**

Fuente: Facultad de Administración, Finanzas e Informática.
Realizado Por: Ernesto Mora

Gráfico Nº 2



Interpretación

En cuanto a la frecuencia que los estudiantes leen sus correos y/o actualiza su bandeja de entrada se obtuvo mediante los siguientes porcentajes los resultados, el 30% afirmo que actualizan sus correos frecuentemente, el 33% dijo que lo realiza poco frecuente, un 21% lo hace rara vez y el 16% dijo que nunca lo hace.

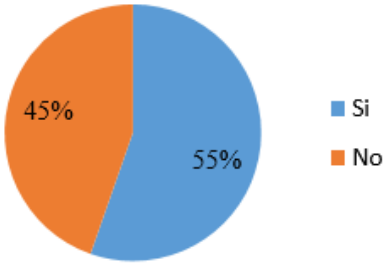
Pregunta # 3 ¿Usted conoce qué es un SPAM y Para qué sirve?

Tabla Nº 3

ALTERNATIVA	FRECUENCIA	PORCENTAJE
Si	227	55%
No	183	45%
TOTAL	**410**	**100%**

Fuente: Facultad de Administración, Finanzas e Informática.
Realizado Por: Ernesto Mora

Gráfico Nº 3



Interpretación

En la siguiente pregunta se comprobó los siguientes resultados acerca del conocimiento sobre SPAM y para qué sirve, en el cual el 55% afirmo que, si conoce sobre correos basura, mientras que el 45% no tiene conocimiento sobre spam.

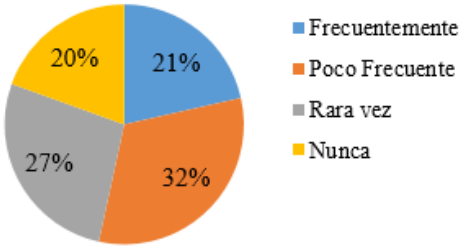
Pregunta # 4 ¿Con qué periodicidad experimenta correos basura o Spam en su e-mail?

Tabla Nº 4

ALTERNATIVA	FRECUENCIA	PORCENTAJE
Frecuentemente	88	21%
Poco Frecuente	131	32 %
Rara vez	111	27%
Nunca	80	20%
TOTAL	**410**	**100%**

Fuente: Facultad de Administración, Finanzas e Informática.
Realizado Por: Ernesto Mora

Gráfico Nº 4



Interpretación

Se reconoce la siguiente pregunta en el cual el mayor porcentaje está en que poco frecuente con un 32% experimenta correos basura o Spam en su e-mail, mientras que un 21% dijo que frecuentemente, el 27% rara vez y el 20% dijo que nunca ha experimentado correos basura en su e-mail.

¿Posee una cuenta de correo institucional o en Gmail?

Si _____ No_____

1. ¿Con qué frecuencia usted lee sus correos y/o actualiza su bandeja de entrada?
   Frecuentemente ____
   Poco Frecuente ____
   Rara vez ____
   Nunca ____
2. ¿Usted conoce qué es un SPAM y Para qué sirve?
   Si_____
   No_____
3. ¿Con qué periodicidad experimenta correos basura o Spam en su e-mail?
   Frecuentemente ____
   Poco Frecuente ____
   Rara vez ____
   Nunca ____
4. ¿En alguna o varias ocasiones usted fue afectado por virus o malwares a través del correo; indique en qué medida lo afectado?
   Muy Afectado ____
   Poco Afectado ____
   Casi Nada ____
   Ninguno ____
5. ¿Ha recibido algún tipo de estos correos?
   Correos de Bancos o Tarjetas ____
   Publicidad no deseada _____
   Cadenas ____
   Otras _____
6. ¿Indique con qué frecuencia actúa efectivamente su sistema de protección ante correos no deseados?
   Frecuentemente ____
   Poco Frecuente ____
   Rara vez ____
   Nunca ____

CONCLUSIONES

La facultad de Administración Finanzas e informática cuenta con las protecciones necesarias contra ataques maliciosos de spam que pueden interferir en la red y a la vez ayuda a evitar los correos no deseados que ocasionen problemas en nuestras cuentas de google, Podemos observar que en los últimos 30 días el control anti Spam funciona correctamente en las cuentas institucionales de la facultad.

De acuerdo a los resultados obtenidos por la realización de las encuestas, se puede evidenciar que la mayoría de la Comunidad Universitaria de la Facultad de Administración, Finanzas e Informática ha sido afectada por virus o malware a través de los correos.

Los Estudiantes deben tener más cuidado cuando traten con cualquier mensaje que los induzca a descargar aplicaciones, cadenas, publicidades, activar vínculos o introducir contraseñas.

Tanto el personal administrado como los docentes

deben de actualizar sus correos y estar atentos a este tipo malware o spam a las que se encuentran vulnerables sus cuentas.

Llegamos a la conclusión de que muchos de los Spam o correo basura que recibes en tu bandeja de entrada sin que lo hayas solicitado normalmente son publicitario, pero también puede llegar a ser peligroso si contiene un enlace o un archivo infectado que lo que único que hacen es que al dar clic sobre el enlace infecta tu equipo.

Para evitar el correo basura, los clientes de correo electrónico actuales han mejorado mucho su detección del correo no deseado y gracias a unos sencillos consejos podemos acabar con el spam o al menos tenerlo tan controlado que apenas nos ocupará unos segundos el mirar la bandeja de spam y borrarla.

RECOMENDACIONES

- Constar con políticas de seguridad informática para así evitar ataques.
- Evitar publicar su dirección de correo en foros, chats, grupos de noticias, etc.
- Evitar conteste a mensajes de correo basura ni abra las páginas Web en el que invitan a conseguir más información o a borrarle de su lista de clientes; con esto sólo se consigue confirmar la existencia de la dirección.
- Que la comunidad universitaria de la Facultad de Administración, Finanzas e Informática, actualice la bandeja de entrada de sus respectivos correos, ya sean institucional o Gmail para evitar ser afectado de alguna manera por correos no deseados.
- Debe ignorar mensajes de correos basura, se le avisa de peligrosos virus o se le indica que los reenvíe a otras personas.
- Nunca descargues archivos adjuntos de e-mails que hayan llegado a la carpeta spam.
- Preferiblemente, tampoco abras correos electrónicos de remitentes desconocidos.
- Utiliza un antivirus que tenga un complemento para integrarse en tu correo electrónico y pueda analizar los correos que recibes.

REFERENCIAS BIBLIOGRÁFICAS

- Cormack, G. S. (2011). MACHINE LEARNING METHODS FOR SPAM E-MAIL CLASSIFICATION. Springer Netherlands.
- Fernández, F. (s.f.). Redactor_CEO. Obtenido de https://www.seoposicionamientoweb.es/que-es-google-safebrowsing/
- Guzella, M.-S. J. (2009). ”A review of machine learning approaches to Spam “. Expert Syst. Appl.
- Juarez, G. (24 de 05 de 2017). Nexolution enabling high performance. Obtenido de http://www.nexolution.com/como-funciona-el-aprendizaje-automatico-machine-learning/
- Koby Crammer, O. D. (2010). Online passive-aggressive. JLMR,.
- Linares, I. (31 de 05 de 2017). El Androide Libre. Obtenido de https://elandroidelibre.elespanol.com/2017/05/gmail-proteccion-contra-phishing-aprendizaje-maquina-inteligencia-artificial.html
- Pacheco, V. G. (2017). Synergic Partners. Obtenido de http://www.synergicpartners.com/una-breve-historia-del-machine-learning/
- Porto, J. P., & Gardey, A. (2016). definicion.de. Obtenido de DEFINICIÓN DE SPAM: https://definicion.de/spam/
- RODRIGUEZ, T. (05 de 04 de 2017). Machine Learning y Deep Learning: cómo entender las claves del presente y futuro de la inteligencia artificial. Obtenido de http://www.Machine%20Learning%20y%20Deep%20Learning%20%20cómo%20entender%20las%20claves%20del%20presente%20y%20futuro%20de%20la%20inteligencia%20artificial.htm
- Yang., S. J. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and.