ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

ITIN
INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

Detección de ataques de Phishing utilizando Procesamiento de Lenguaje Natural y Modelo Oculto de Markov

Autores:

Monteros González, Kevin Joel

Molina Salavarría, Jhoseph Alberto

Director:

Ing. Benavides Astudillo, Diego Eduardo, Mgtr

“Si crees que la tecnología puede solventar tus problemas de seguridad, entonces no entiendes los problemas y no entiendes de tecnología.”

Bruce Schneier

Agenda

Introducción	3 min
Problemática	3 min
Justificación	3 min
Objetivos	2 min
Marco Teórico	7 min
Metodología	7 min
Resultados y Discusión	10 min
Conclusiones, Trabajo Futuro y Recomendaciones	5 min

Introducción

La Ingeniería Social continúa siendo el método de propagación de ataques informáticos más utilizado por los creadores de malware, quienes aprovechan las ventajas de cualquier medio de comunicación para engañar a los usuarios y lograr sus objetivos maliciosos.

Problemática

El uso del correo electrónico trae consigo amenazas y riesgos que surgen de vulnerabilidades explotadas por atacantes que buscan obtener acceso ilegal a diversa información y causar daño dentro de una organización o directamente a un individuo.

En los últimos años los ataques de Ingeniería Social han incrementado descontroladamente, ya que este tipo de seguridad ya no depende de equipos de alto rendimiento o buenos sistemas, sino que depende en gran medida del usuario y de su capacidad para no entrar en el juego de este tipo de ciberdelincuentes.

Justificación

Aunque la efectividad de los softwares, firewalls y medidas de seguridad basadas en hardware para contrarrestar delitos informáticos se han incrementado a la par con los tipos de ciber ataques, la exposición por parte del usuario, que es el eslabón más débil de la seguridad informática ante ataques de Ingeniería Social, no ha cambiado; esto se debe a que no existen mecanismos eficaces para anticipar las acciones del usuario, por lo cual, quedan expuestos y vulnerables.

Objetivos

Objetivo general

Realizar un estudio comparativo entre el Hidden Markov Model (HMM) y los algoritmos tradicionales de Machine Learning (ML) para la detección de correos electrónicos Phishing.

Objetivos específicos

- Revisión de la literatura.
- Obtención y pre procesamiento del Dataset a ser utilizado.
- Implementación de los algoritmos de Machine Learning.
- Implementación del algoritmo de HMM.
- Análisis comparativo y discusión.

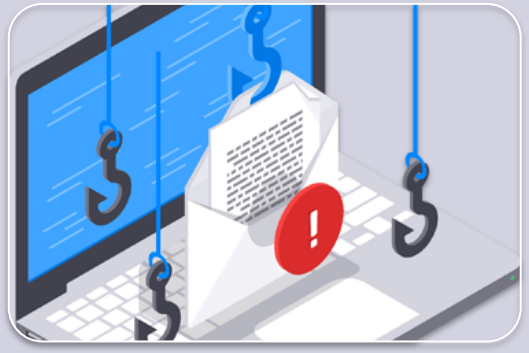
Marco teórico

Ingeniería social

Es una técnica de manipulación que explota el error humano para obtener información privada, acceso u objetos de valor.

Spam

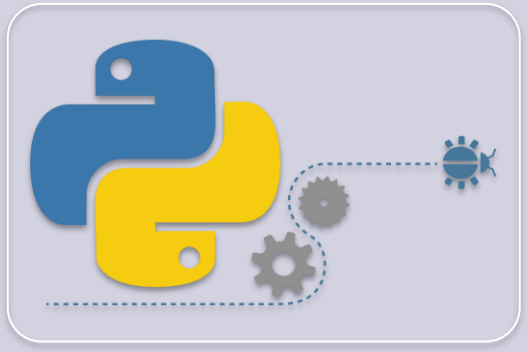
Es cualquier tipo de comunicación digital no deseada y no solicitada que se envía en masa



Phishing

En un tipo de ciberataque más sencillo y, al mismo tiempo, el más peligroso y eficaz. Esto se debe a que ataca al ordenador más vulnerable y poderoso del planeta: la mente humana.

Marco teórico



Machine Learning (ML)

Disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones y elaborar predicciones.

Natural Language Processing (NLP)

Disciplina de estudio centrada en cómo los ordenadores entienden el lenguaje humano, lo interpretan y procesan.

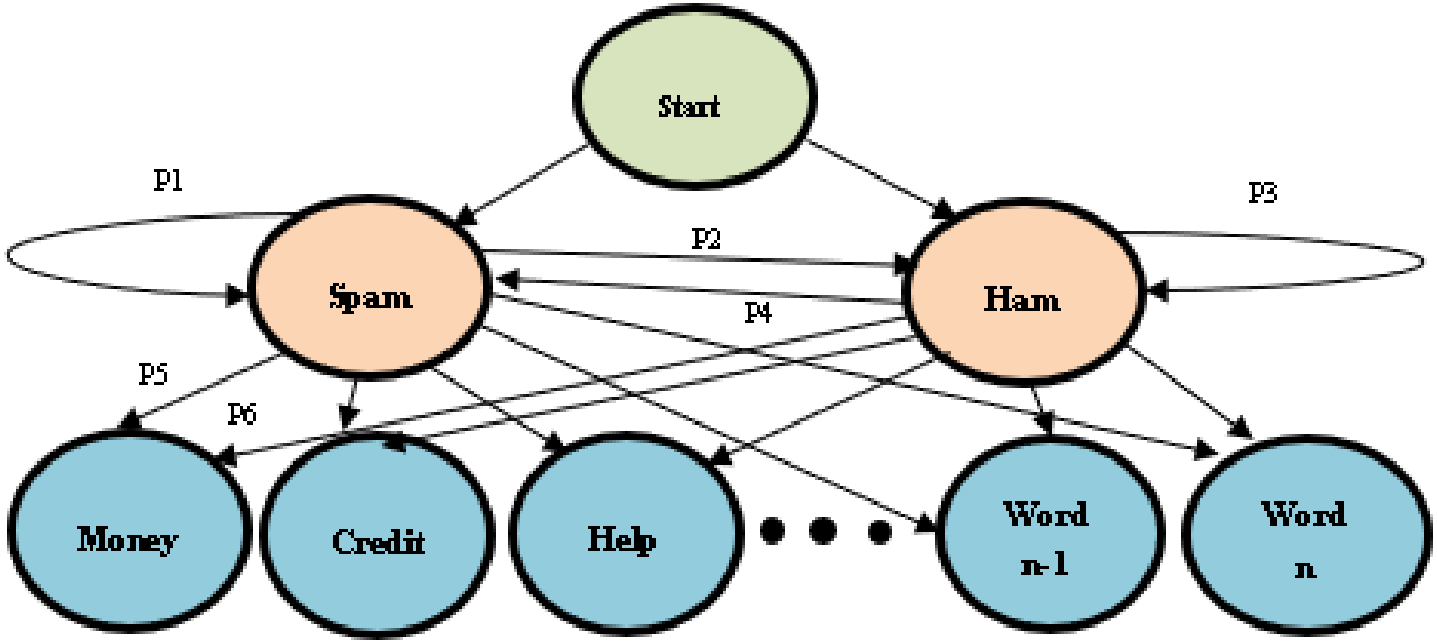
NLTK

Conjunto de bibliotecas y programas para el procesamiento del lenguaje natural simbólico y estadísticos para el lenguaje de programación Python.

Marco teórico

Hidden Markov Model (HMM)

Es una cadena de Markov donde cada estado, que no es observable y está asociado con una distribución de probabilidad, genera una observación que sí es observable.



Metodología

Revisión sistemática de la literatura (1 - 3)

Metodología de Bárbara Kitchenham (B. Kitchenham, 2009)

Planificar la revisión

Conducir la revisión

Documentar la revisión

Se recopilaron 24 artículos del filtro aplicado en distintas bases de datos científicas.

Metodología

Revisión sistemática de la literatura (2 - 3)

Criterios de clasificación:

De calidad

- Scopus
- Web of Science
- Escritos en inglés

De inclusión / exclusión

- Artículos con resultados

- Aplicación de NLP

- Técnica de IA

18 artículos

Metodología

Revisión sistemática de la literatura (3 - 3)

Selección:

- Lightweight URL-based phishing detection using natural language processing transformers for mobile devices **(Haynes, 2021)**
- Machine Learning and Deep Learning for Phishing Email Classification using One-Hot Encoding **(Bagui, 2021)**
- A hybrid DNN-LSTM model for detecting phishing URLs **(Ozcan, 2021)**

Nota: Ninguno de los artículos seleccionados utilizan HMM.

Metodología

Tipo de investigación

La investigación presentada es de tipo cuantitativa y descriptiva

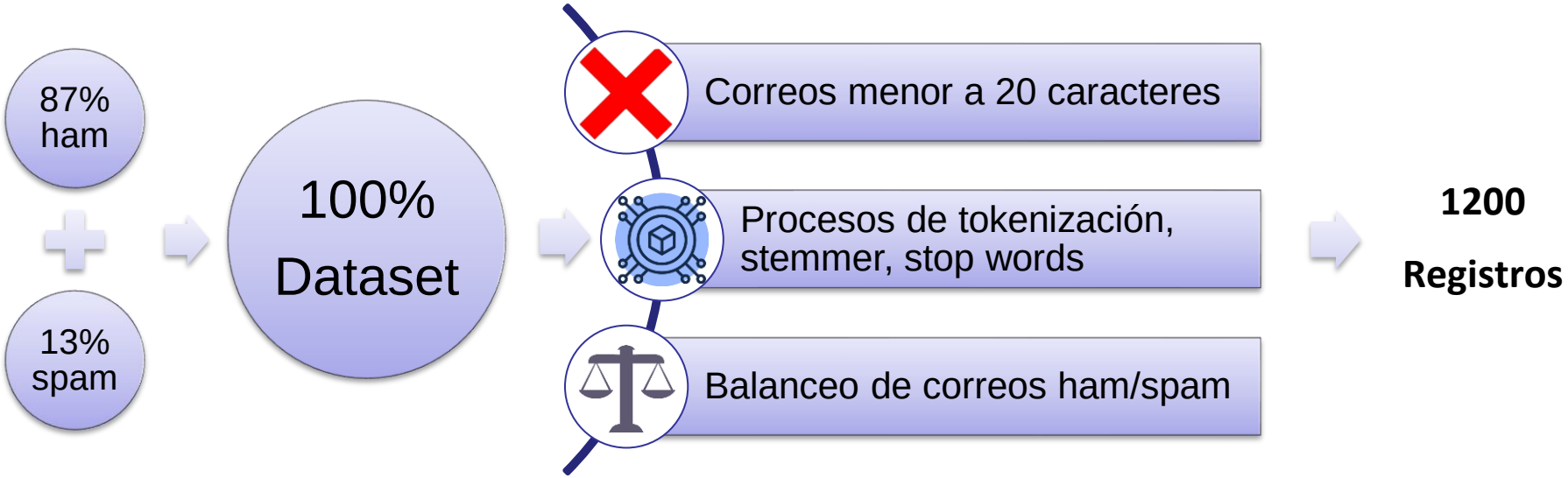
Maneja valores concretos en cuanto a exactitud y precisión

Se presenta un análisis a los datos obtenidos en la ejecución de los algoritmos

Metodología

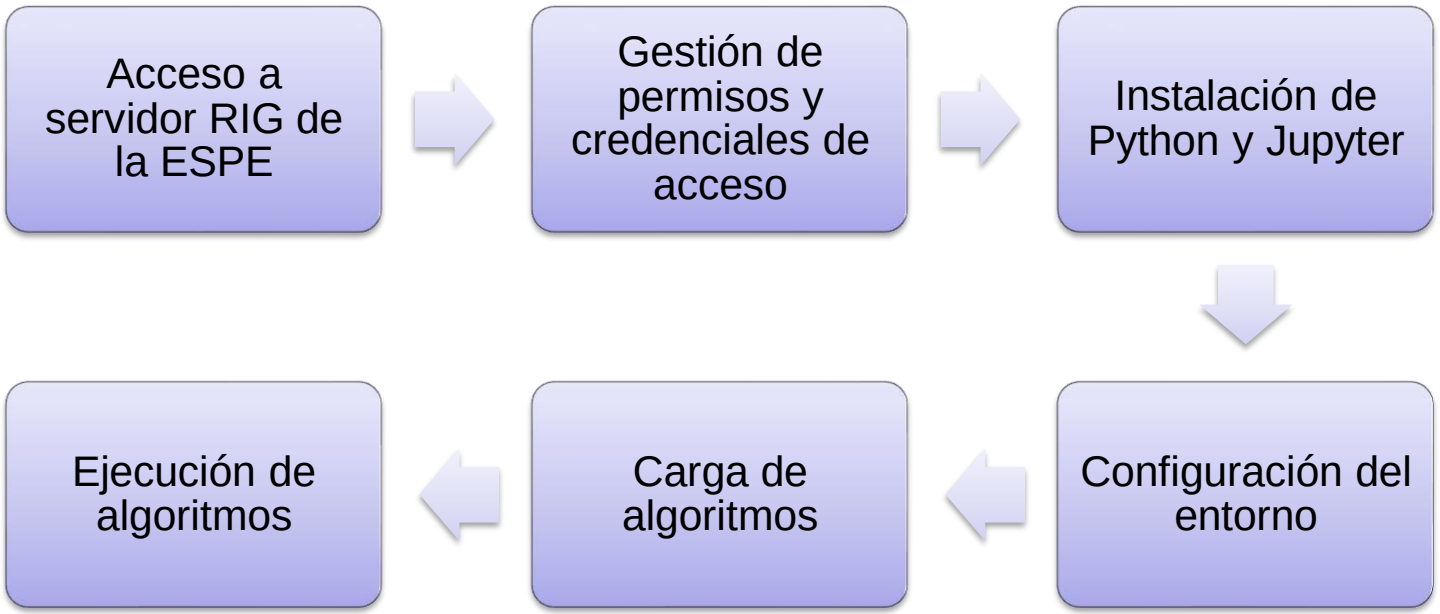
Dataset utilizado

Dataset de 5172 registros obtenido de (Dhakad, 2018) “Phishtank”



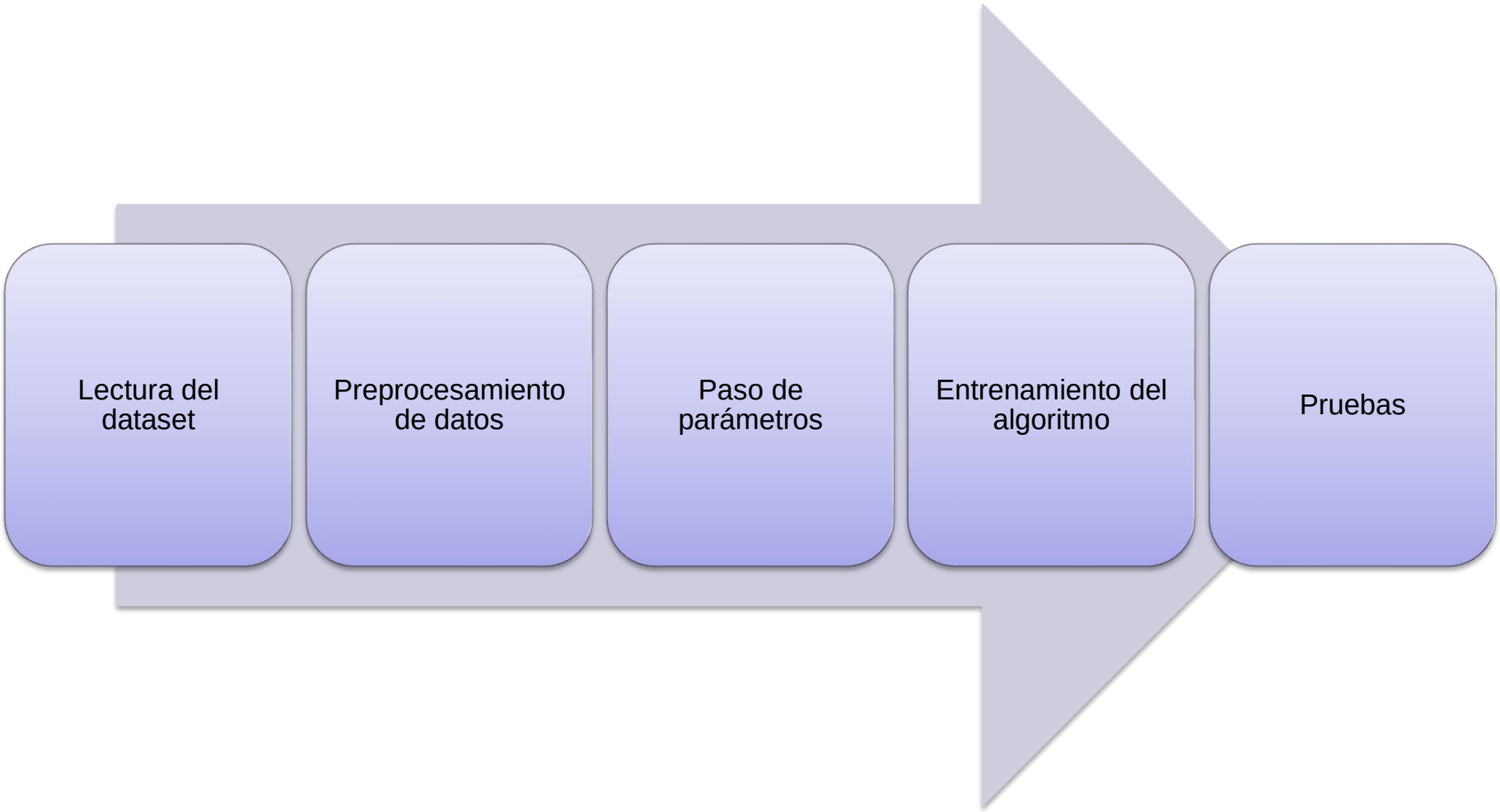
Metodología

Procedimiento realizado



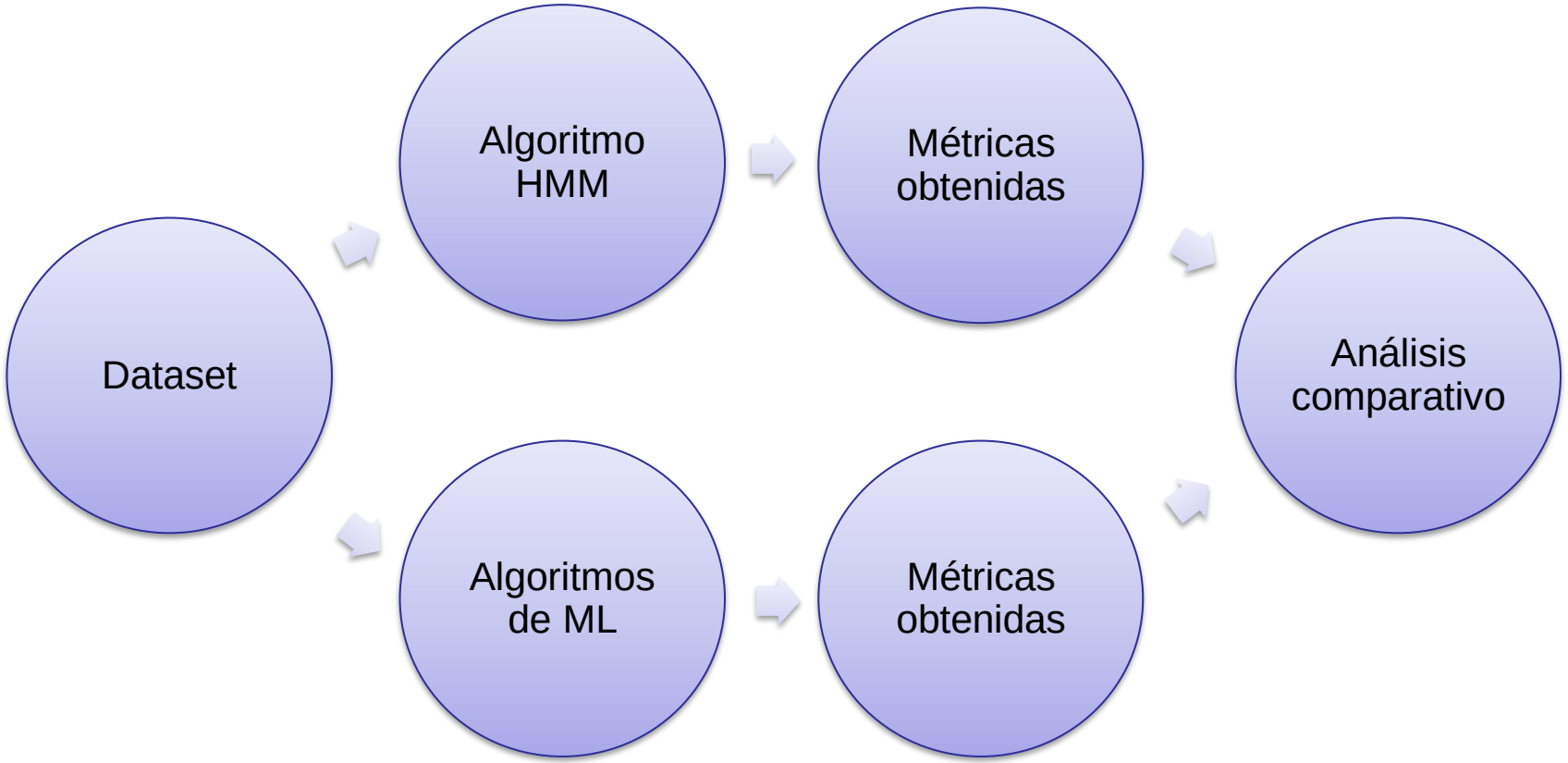
Metodología

Funcionamiento de los algoritmos



Metodología

Procedimiento de análisis



Resultados

Resultados de la revisión sistemática de la literatura

Nro	Paper	Precisión	Nro de Registros
1	(Haynes, 2021)	96.00%	21,910
2	(Bagui et al, 2021)	96.34%	18,366
3	(Ozcan, et al, 2021)	99.21%	99,575
4	(al X. X., 2021)	95.60%	80,033
5	(Alsufyani, 2021)	90.00%	10,306
6	(Junnarkar, 2021)	94.64%	110
7	(Indrasiri et al, 2021)	98.27%	173,575
8	(Yaseen, 2021)	98.67%	11,297
9	(Sirigineedi, 2020)	96.60%	73,575
10	(Kumar, 2020)	98.00%	1,705
11	(Sahingoz, 2019)	97.00%	73,575
12	(Verma, 2019)	80.00%	3,865
13	(Thakur, 2018)	87.00%	286,000
14	(Buber, 2018)	97.20%	7,357
15	(Buber, Dırı, & Sahingoz, 2017)	89.90%	10,572
16	(R. Verma, 2012)	97.00%	2,000
17	(Ona et al, 2019)	93.90%	3,000
18	(Espinoza, 2019)	96.77%	1,325

Resultados

Resultados de la ejecución del Algoritmo HMM

In [70]:
```
import pandas as pd
data1 = pd.read_csv("prueba5000.csv", lineterminator='\n')
data1
data2=data1.head(100)
```

In [120]:
```
cm = confusion_matrix(y_test, y_pred)
Accuracy_Score = accuracy_score(y_test, y_pred)
Precision=precision_score(y_test, y_pred, average='weighted')

print('An algorithm to detect new Spam attacks with Hidden Markov Model vs Machine Learning')
print('By Kevin Monteros and Jhoseph Molina.')
print('\n')
print('Confusion Matrix')
print(cm)
print('\n')
print('Accuracy Score')
print (round (Accuracy_Score,2))
print('\n')

```

```
An algorithm to detect new Spam attacks with Hidden Markov Model vs Machine Learning
By Kevin Monteros and Jhoseph Molina.


Confusion Matrix
[[188  70]
 [  1  42]]


Accuracy Score
0.76
```

Resultados

Resultados de la ejecución de los algoritmos de ML

```
In [60]:  1  pred_scores_word_vectors

Out[60]: [('DT', [0.96]),
          ('SVC', [0.9533333333333334]),
          ('KN', [0.8933333333333333]),
          ('NB', [0.98]),
          ('RF', [0.9533333333333334]),
          ('LR', [0.9])]
```

Discusión

En base a los resultados obtenidos, se puede observar que el algoritmo HMM da una precisión del 76%, es decir, es inferior a cualquiera de los algoritmos de Machine Learning que implementamos (DT = 96%, SVC = 95,53%, KN = 89,33%, NB = 98%, RF = 95,33%, y LR = 90%).

A primera vista, se puede observar que HMM da resultados más bajos que ML; sin embargo, hay que tener en cuenta que también en ML se realizó un análisis semántico utilizando NLTK.

Conclusiones, Trabajo Futuro y Recomendaciones

Conclusiones

- En este trabajo primero se realizó una revisión de la literatura, en la que se exponen las principales características obtenidas de las soluciones orientadas a resolver problemas de Spam utilizando NLP, en segundo lugar, se apertura una nueva línea de investigación que propone incrementar el uso de algoritmos HMM en la detección de correos Spam fraudulentos.

- El tiempo de procesamiento de datos en el caso del algoritmo HMM es muy alto, esto se debe a que analiza de manera profunda cada palabra y realiza predicciones en base a dicho análisis. Los algoritmos de ML manejan un tiempo de respuesta bastante corto, debido a que el funcionamiento de estos, no se basa en un análisis semántico profundo.

Conclusiones, Trabajo Futuro y Recomendaciones

Trabajo futuro

- Como trabajo futuro, pretendemos desarrollar un algoritmo que ofrezca mayor precisión, en el que el texto de entrada sea analizado primero por el algoritmo HMM y luego introducido en un conjunto de algoritmos de ML, concretamente DL.

- Otra propuesta, es la de implementar alguna solución de software que permita la implementación de ML, DL o HMM, a fin, de que dicho programa pueda detectar un correo de Spam antes de ser abierto. De esta manera, se podría advertir a tiempo al usuario sobre un posible ataque de Ingeniería Social.

Conclusiones, Trabajo Futuro y Recomendaciones

Recomendaciones

- En base a la investigación realizada, hasta el momento se siguen recomendando las alternativas que ofrece ML en la detección de correos de Spam debido a su alto porcentaje de exactitud. Sin embargo, HMM mantiene una línea de investigación latente y no explorada, a diferencia de muchos algoritmos que con el pasar del tiempo han llegado a quedar obsoletos al trabajar de manera autónoma.

- Se recomienda contar con equipos de cómputo con la suficiente capacidad para realizar ejecución de algoritmos y procesamiento de datos en gran cantidad y a buen tiempo, ya que para realizar la ejecución del algoritmo HMM se contó con un servidor de excelentes características, aun así, no fue suficiente para realizar las debidas pruebas y obtener resultados en menos de 24 horas.

GRACIAS!

“El mayor riesgo es no correr ningún riesgo”
Marck Zuckerberg

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA