MODELO PREDICTIVO PARA IDENTIFICAR DELITOS DE ACOSO EN LA RED SOCIAL FACEBOOK APLICANDO BIG DATA

Leyna Roxana Salinas Veyzaga, Ph.D.
leynasud@gmail.com
Ingeniería Informática
Universidad Nacional "Siglo XX"
Llallagua, Bolivia

Resumen- *El uso masivo de las redes sociales ha hecho que los delincuentes puedan usar estas tecnologías con la finalidad de cometer sus crímenes, uno de ellos es el acoso en línea. El objetivo es diseñar un modelo predictivo basado en Big Data para identificar delitos de acoso en la red social Facebook. Este modelo toma como dataset información extraída de los comentarios de publicaciones en Facebook, para utilizar un algoritmo de Machine Learning, teniendo como resultado un modelo de predicción que pueda presentar patrones de acoso y no acoso, los cuales se manifiestan en agresión verbal grave, como insultos, ataques racistas, ataques homofóbicos, etc.*

Palabras clave- *Big Data, Dataset, FacePager, Machine Learning, Matriz de confusión, Naive Bayes, Redes Sociales.*

Abstract- *The massive use of social networks means that criminals can use these technologies in order to commit their crimes, one of them is online harassment. The objective is to design a predictive model based on Big Data to identify harassment crimes on the Internet. social Facebook. This model takes as a data set information extracted from the comments of posts on Facebook, to use a Machine Learning algorithm, resulting in a prediction model that can present patterns of harassment and non-harassment, which manifest themselves in serious verbal aggression, such as insults, racist attacks, homophobic attacks, etc.*

Keywords- *Big Data, Confusion Matrix, Dataset, FacePager, MachinE Learning, Naive Bayes, Social Networks.*

1. INTRODUCCIÓN

Con una primera observación los entornos sociales, las noticias, y multiplicidad de tareas cotidianas, se puede inferir que Internet tiene una importancia central que organiza el sistema de información en las sociedades actuales, y por tanto es clave realizar estudios de fenómenos en la red que posibiliten conocer el nuevo orden, funcionamiento y comprensión de una multiplicidad de fenómenos socio técnicos relacionadas a la red (Gómez & Farrera, 2019).

Actualmente para los criminales resulta mucho más sencillo actuar de manera digital que presencial, ya que ellos pueden operar creando perfiles falsos con el fin de atacar a sus víctimas.

El acoso cibernético mediante las redes sociales es un tema que tiene un carácter innovador y que pretende disminuir los delitos por intermedio de las redes sociales, ya que las mismas son aplicaciones de esta última generación.

La cantidad de datos que se están moviendo y generando en las redes sociales es demasiado grande, ahí es donde ingresa el término de Big Data ya que permite analizar la información generada con la finalidad de predecir comportamientos futuros y tomar decisiones.

No existen investigaciones de predicciones de acoso en línea mediante técnicas de Machine Learning en la red social Facebook en Bolivia.

Así se puede apreciar, que se necesita un procedimiento para identificar delitos de acoso en la red social Facebook aplicando Big Data.

El objetivo general de la presente investigación es diseñar un modelo predictivo para identificar patrones de comportamiento de personas que cometen delitos de acoso en la red social Facebook aplicando Big Data.

2.- MATERIALES Y METODOLOGÍA

Con el fin que persigue la investigación será una Investigación Exploratoria y corresponde a un tipo transformador utilizando el método Inductivo- deductivo. Tomando como población a personas exclusivamente que sufrieron acoso en la red social Facebook en Bolivia por medio de comentarios ofensivos y degradantes.

El sitio Guía anti-acoso, muestra la incidencia de acoso digital en Bolivia y brinda algunos consejos para actuar en caso de sufrir acoso digital en Facebook.

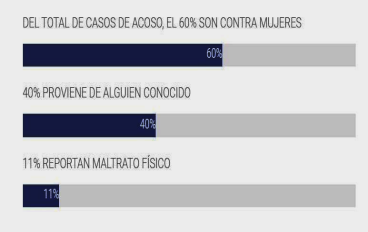


Figura 1. Incidencia de acoso digital en Bolivia
Fuente: (internetbolivia, 2013)

Nota. Representa el porcentaje de incidencia del acoso en Bolivia. Tomado de https://internetbolivia.org/8m/

Al mismo tiempo este sitio web muestra los perfiles que son más vulnerables, los cuales son:

- Mujeres que se encuentran en una relación íntima con parejas de antecedentes violentos.
- Mujeres con perfiles públicos (periodistas, activistas, empresarias).
- Sobrevivientes de violencia física o sexual.

Según el sitio Sistema para la prevención de ciberacoso para el idioma español, el cual contiene recursos que son parte del Proyecto "Detección presuntiva de ciberacoso en redes sociales basado en el análisis de sentimientos".

El sitio contiene un corpus de términos con 13.430 términos únicos con sus respectivos pesos. La siguiente tabla muestra algunos términos con su peso respectivo.

N	Comentario	Etiqueta
1	horita que mi perro te limpie ese culo cagado puta cabrona y luego te vas a abotonar con mi perro mendiga perra callejera	1.0
2	Los que hacen chistes de lo que le ocurrió a esa señorita, nada bien, absolutamente nada bien en casa, verdad?	0.0
3	Como siempre hermosa y ricota mi amor siempre haciéndome babear con tu hermoso culito mi reina	1.0
4	El machismo en el área de la salud empieza cuando crees que por ser mujer soy la enfermera y no la doctora.	-1.0

Tabla 1. Algunos términos del corpus de Ciberacoso
Fuente: Elaboración propia

Nota. Representa algunos términos del corpus con sus respectivos pesos. Tomado de https://cloudcomputing.ups.edu.ec/SCPSystem/faces/acerca_corpus.xhtml

Se extrajo comentarios de publicaciones de perfiles en base a una búsqueda mediante palabra clave, llegando a un total de 1196 comentarios para la construcción del conjunto de datos.

La presente investigación propone un modelo predictivo que utiliza técnicas de análisis de datos en Big Data, estás se enmarcan dentro de los procesos que comprenden la recolección, depuración, tratamiento, modelado y estudio de datos encaminados a la obtención de conclusiones útiles en este caso será para identificar patrones de comportamiento de personas que cometen delitos de acoso en la red social Facebook.

1) Recopilación de Datos, por la particularidad que presenta la investigación se recurrirá a la técnica Web Scraping. Se utilizó el software Facepager, que es un software automatizado para recolectar los datos de las redes sociales como Facebook, esta herramienta permite la obtención de los comentarios de un post de Facebook para que la información extraída pueda ser analizada. Se etiquetó con un 0 para ver si no existe acoso en un comentario de un post publicado; y un 1 para ver si existe acoso en un comentario de un post publicado, también se consideró los comentarios que son neutros y fueron etiquetados con un -1.

La siguiente tabla muestra algunos ejemplos que son acoso (1), no acosó (0) y neutro (-1)

Peso	Término
0.0162000575289	Puta
0.0125960602248	Quiero
0.0103446856503	Gente
0.0103331944324	Vida
0.0099205041536	Mierda
0.00968068833262	Amigo
0.00941831146553	Amo
0.00858364831772	Culo
0.00831919609048	Perro
0.00803216401749	Siento
0.00800279242667	Amiga
0.00800074496117	Madre
0.0069177313417	Ricota
0.00687980014533	Perra

Tabla 2. Ejemplos etiquetados de algunos comentarios extraídos
Fuente: Elaboración propia
Nota. Representa algunos comentarios etiquetados. Elaboración propia.

2) Preprocesamiento de datos

Es muy importante preprocesar los datos extraídos de los comentarios de una publicación de Facebook, generalmente implica varias etapas con la finalidad de limpiar y preparar los datos para un análisis posterior. A continuación, se describe las técnicas para el preprocesamiento de datos: limpieza de datos; los comentarios extraídos podrían contener, caracteres especiales, o formatos no deseados como etiquetas HTML, para ellos se escribieron funciones para convertir los textos a minúsculas, eliminar caracteres especiales, eliminar las palabras vacías; Tokenización, se utilizó esta técnica, la cual permite dividir el texto en unidades más pequeñas, como palabras individuales o frases, para facilitar el análisis; finalmente, cada comentario se codifica en un vector numérico

que representa la frecuencia de cada palabra en el comentario. Se utilizó la técnica de conteo de términos (Count Vectorizer)

3) Análisis exploratorio

Antes de proceder con un análisis más profundo, es útil realizar visualizaciones preliminares como histogramas, nubes de palabras (word clouds) o gráficos de dispersión para entender mejor la distribución y patrones de los datos de los comentarios de Facebook.

La siguiente figura muestra la nube de palabras de los comentarios etiquetados como acoso (1)

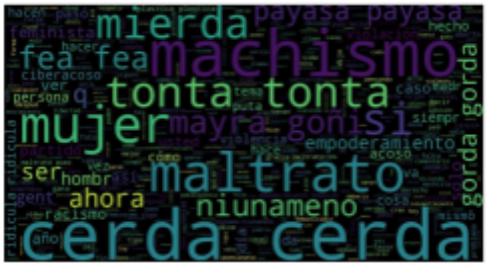
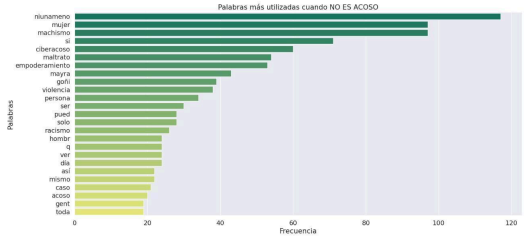


Figura 1. Nube de Palabras de los comentarios etiquetados como Acoso.
Fuente: Elaboración propia

Nota. Representa la nube de palabras de los comentarios etiquetados como acoso. Elaboración propia.

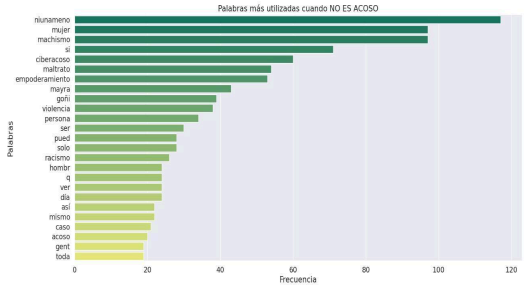
La gráfica 1 muestra la distribución de las Palabras más utilizadas cuando el comentario es etiquetado como acoso.



Gráfica 1. Distribución de palabras más utilizadas cuando es acoso.
Fuente: Elaboración propia

Nota. Representa la distribución de palabras más utilizadas cuando es acoso.

Por otro lado, se tiene la distribución de las Palabras más utilizadas cuando No es acoso



Gráfica 2. Distribución de las Palabras más utilizadas cuando No es acoso
Fuente: Elaboración propia

Nota. Representa la distribución de palabras más utilizadas cuando no es acoso.

4) Construcción del modelo predictivo, en la construcción del modelo se tiene las siguientes etapas:

Selección de algoritmo, en esta etapa se utilizan técnicas avanzadas de Big Data como el Machine Learning, se utilizó el Algoritmo Naives Bayes para el entrenamiento. Tomando la recopilación de términos únicos, teniendo como entrada el conjunto de palabras que aparecen en el comentario y produciendo como salida una lista de términos únicos.

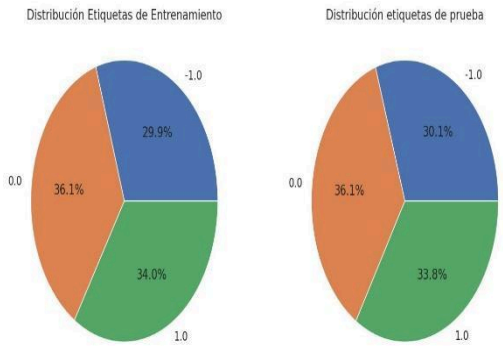
Entrenamiento del modelo de predicción, con los datos codificados y etiquetados, se procede a entrenar el modelo con el algoritmo seleccionado.

Para el entrenamiento se requieren los siguientes parámetros de entrada: se considera un 75 % para datos de entrenamiento y un 25 % para datos de prueba.

Comentarios de entrenamiento: 895

Comentarios de test: 299

La gráfica 3 muestra la distribución de las etiquetas de entrenamiento y de prueba.



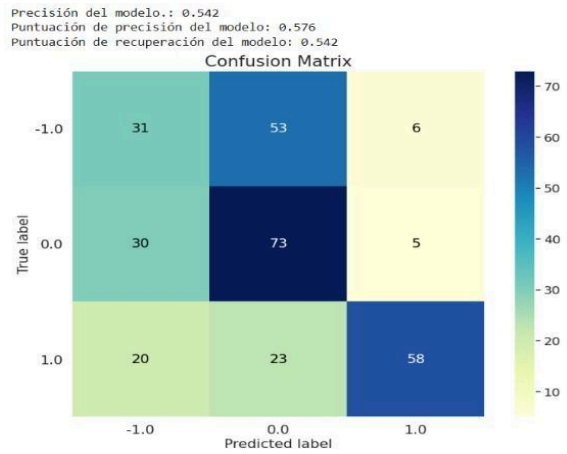
Gráfica 3. Distribución de las etiquetas de entrenamiento y de prueba.
Fuente: Elaboración propia

Nota. Representa la distribución de etiquetas de entrenamiento y de prueba. Elaboración propia.

Evaluación del modelo, para la evaluación del modelo se utilizó la matriz de confusión, esta métrica permite evaluar el rendimiento del modelo. Una vez que el modelo está entrenado, se debe evaluar su rendimiento utilizando un conjunto de datos de prueba. Esto proporciona una idea de cuán bien generalizada está el modelo a nuevos comentarios no vistos.

3.- RESULTADOS

Una vez que el modelo fue entrenado con los 3 tipos de comentarios etiquetados con los datos separados tanto para las pruebas como para las predicciones, se tienen los siguientes resultados reflejados en la gráfica 4.



Gráfica 4. Matriz de confusión.
Fuente: Elaboración propia

Nota. Representa la matriz de confusión del modelo. Elaboración propia

En la matriz de confusión se puede evidenciar de los 299 comentarios de pruebas:

El modelo predijo correctamente:

58 comentarios que es de acoso.

5 comentarios de No son de acoso

6 comentarios que son Neutros.

El modelo alcanza una precisión del 80-90% en la detección de acoso.

4.- CONCLUSIONES

En los resultados de la investigación se puede apreciar que las técnicas de Big data ayudan a predecir acoso cibernético y estas mismas son utilizadas para diferentes fines, tal es el caso de del trabajo de Maestría referido al uso de las redes Sociales intitulado: Análisis predictivo en Twitter para detectar patrones de personas con tendencia Hacktivista aplicando Big Data, Machine Learning y Deep Learning. En esta investigación, se pretende desarrollar un modelo que permita la identificación de vocabulario hacktivista, a través de la combinación de técnicas de data mining y algoritmos de machine learning y Deep learning, el resultado será la implementación de un modelo con bastante precisión en la identificación y clasificación del vocabulario que refiere a Hacktivismo.

REFERENCIAS

Gómez, D.M., & Farrera, R.A. (2019). El hacktivismo e Internet como territorio en disputa. Una mirada desde los marcos de acción colectiva. Estudios Políticos.

Joyanes Aguilar, L. (2013). *Big Data. Análisis de grandes volúmenes de datos en organizaciones*. México: Alfaomega.

Ramos-Vidal, I. (2015). Análisis de redes sociales: una herramienta efectiva para evaluar coaliciones comunitarias. *Revista de Salud Pública, 17(3)*, 323-336.

Raschka, S. (2019). *Python Machine Learning "Aprendizaje automático y aprendizaje profundo con Python, scikit-learn y TensorFlow"* (2° Edición ed.). Barcelona: Marcombo.

Rodríguez Devesa, J. M. (1974). Derecho Penal Español. *Parte General*, 75-75. Obtenido de https://dialnet.unirioja.es/servlet/articulo?codigo=8652894.

Rueda Buste, J. L. (2023). El uso de redes sociales y su incidencia en el delito de acoso sexual en adolescentes. *Artículo Científico de Abogado (a) de los Tribunales de la República*, 28.

Sanchez, L. (2003). Análisis de redes sociales: O cómo representar las estructuras sociales subyacentes. *Unidad de Políticas Comparadas*. Obtenido de https://www.researchgate.net/publication/260184831_Analisis_de_redes_socialeO_como_representar_las_estructuras_sociales_subyacentes