

Análisis de la incidencia delictiva del fuero común mediante la red social Twitter

Rosa G. Hernández Cuacua, Sandra D. Rosas Ortega,
M. Beatriz Bernábe Loranca

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación,
México

{gabyrocklml1, daana.orteeg,
beatriz.bernabe}@gmail.com

Resumen. En este trabajo se ha hecho uso de técnicas de Procesamiento de Lenguaje Natural PLN, para describir la situación actual en México respecto a la inseguridad (incidencia delictiva del fuero común). El estudio inicia con descubrir opiniones de usuarios en la red social Twitter. Para este propósito, se parte de extraer información aleatoria usando palabras representativas del problema etiquetándolas como hashtag. El desarrollo del trabajo se centra en la creación de un diccionario que permita contabilizar las palabras clave de incidencia delictiva para proceder a representar gráficamente la información semántica de las perspectivas de los asistentes a Twitter sobre delincuencia generalizada. Una vez concentrada la información que ha sido depurada y estructurada, se procesa el conjunto de información con un análisis de sentimientos. A partir de este punto, ha sido posible deducir mecanismos de reacción frente a un problema de violencia. La respuesta denominada reacción tiene al menos 3 interpretaciones en distintos niveles: miedo, terror o pánico.

Palabras clave: Análisis de sentimientos, inseguridad, procesamiento del lenguaje natural, redes sociales.

Analysis of Crime Incidence of Civil Courts through the Social Network Twitter

Abstract. In this work, Natural Language Processing techniques have been used and applied to describe the current situation in México regarding insecurity (crime incidence of civil courts). The study begins discovering the user's reviews on the social network Twitter. For this purpose, random information is extracted using representative words of the problem by tagging them as hashtag. The development of this work is focused in the creation of a dictionary, which allows counting the key words of crime incidence to proceed with representing graphically the semantic information of the different user's perspectives about the generalized delinquency. Once that the information (that has been "cleaned" and structured) is concentrated, it is processed with Sentiment Analysis. From this point, it is possible to deduce mechanisms of reaction face a violence

problem. The so-called reaction response has at least three interpretations in different levels: fear, horror or panic.

Keywords: Sentiment analysis, insecurity, natural language processing, social network.

1. Introducción

La inseguridad social es un tema muy importante en diversos países en vías de desarrollo, lo cual se traduce en un reto a resolver. Tan sólo en México la tasa de inseguridad es de las más altas reflejándose en una de las mayores preocupaciones de los habitantes de cada estado. Se asume que en este problema destaca la angustia que impacta en la integridad física, emocional y la economía.

Existen dos tipos de delincuencia o actos delictivos que son relevantes en estudios de esta magnitud, sin embargo, en este artículo se aborda la delincuencia menor o común. Esta delincuencia afecta a la mayoría de la población donde no importa la edad, nivel de educación y/o nivel socioeconómico. Esta praxis delictiva da lugar a otras complicaciones dado que sucede diariamente con alto riesgo e impacta en las actividades constantes del riesgo implícito de ser atacado.

Actualmente los actos delictivos suelen presentarse con mayor frecuencia y afectan a casi todos los grupos sociales. En México, estos datos se reportan y registran a los medios de atención a víctimas y consecuentemente, distintas instituciones u organismos gubernamentales organizan esas denuncias con el fin de describir estadísticamente los denominados “focos rojos” de asalto y establecer programas de atención policiaca o de vigilancia.

Desde el enfoque computacional, este tipo de información también puede ser obtenida por otras herramientas de extracción o descubrimiento de información. En este punto, dentro de la Inteligencia Artificial, se encuentra el Procesamiento del Lenguaje Natural (PLN), el cual es conveniente para analizar problemas de tal magnitud. Para el caso que se expone en este artículo, se han identificado las posibles reacciones del problema relacionado con la delincuencia en la red social Twitter. Con el propósito de obtener información asociada con la delincuencia menor, se desarrolló un programa en Python para descargar tweets que identifican palabras y hashtags sobre el problema bajo estudio. Una vez extraída la información, esta es procesada para la creación de diccionarios y análisis de sentimientos.

El trabajo se encuentra estructurado como sigue; la presente introducción como sección 1. La sección 2 se encarga de exponer el marco teórico con los conceptos más relevantes en los últimos cinco años, los antecedentes y estado del arte. En la sección 3 se describen las estrategias para alcanzar el diseño de un algoritmo capaz de soportar descargas de gran tamaño de tweets para crear diccionarios. Esta tarea ha sido posible por la aplicación de un programa desarrollado en Python para PLN. La sección 4 trata el análisis de sentimientos. Y finalmente se presentan las conclusiones y el trabajo futuro.

1.1. Objetivo

El propósito del presente trabajo consiste en recolectar y analizar información sobre la inseguridad en diferentes estados de la república. Este trabajo responde a la necesidad de analizar y atender el problema polarizado de actos delictivos los cuales son cometidos diariamente y que se ha incrementado considerablemente. De esta manera es importante contar con información correcta e insesgada para sensibilizar a los ciudadanos sobre la situación actual que sufre la población en cuanto a delincuencia.

Se entiende que información correcta en este problema, es la generada por el uso de herramientas de PLN en la red social Twitter para obtener datos adecuados sobre la delincuencia. Una vez concentrados datos e información en diccionarios que revelen el problema, es de utilidad su representación en estadística descriptiva para ser comprendida por una comunidad considerable. Posteriormente un análisis de sentimientos se hace necesario para determinar el grado de ansiedad y fragilidad de las personas sobre la delincuencia.

2. Marco teórico

La inteligencia Artificial (IA) es un área reciente que surge alrededor de hace 60 años. Definir un único concepto de Inteligencia Artificial, la cual ha ido madurando a lo largo de los años, no es simple. Sin embargo, en un contexto informal, la IA se expresa como la inteligencia basada en el nivel de conocimientos individuales, y en la habilidad de manipular y reformular apropiadamente los conocimientos de acuerdo con los datos que se proporcionan como requerimientos para resolver algún problema [1]. En términos prácticos se dice que la IA es un campo de estudio que busca explicar y emular el comportamiento inteligente con recursos computacionales [2]. Desde la perspectiva de la informática, la IA trata del diseño de computadoras inteligentes que exhiban las características que se asocian con la inteligencia en el comportamiento humano. Para fines pragmáticos, la IA es una rama de las ciencias computacionales encargada del estudio de la inteligencia con elementos artificiales para llegar a diseñar o plantear sistemas que puedan ser capaces de realizar tareas que hasta hace años sólo la inteligencia humana era capaz de construir manualmente [3].

Dentro de la IA destaca el Procesamiento del Lenguaje Natural (PLN) y se entiende como el uso de los lenguajes naturales para la comunicación de los individuos, ya sea de manera oral, escrita o signada. En este escenario donde el presente artículo se ubica en la extracción de información a través de los lenguajes naturales, explorando instrumentos computacionales que ayuden a describir el problema de incidencia delictiva desde el enfoque de diccionarios y sentimientos en Twitter.

El análisis de sentimientos en textos es la identificación y extracción de información subjetiva; también llamado “minería de opiniones”. Este proceso generalmente involucra el uso de herramientas de PLN y software de análisis de textos para la automatización de estos [4]. El PLN también incluye tratamientos diversos de las redes sociales en donde han tenido gran impacto.

Las redes sociales son lugares en Internet donde las personas publican y comparten todo tipo de información, personal y profesional con terceras personas, conocidos y absolutos desconocidos [5].

En particular, Twitter es una herramienta que comunica principalmente por medio de mensajes cortos, llamados “tweets” que son acompañados de palabras clave, “hashtags” y permiten la agrupación de la información, generalmente suelen tener noticias, opiniones y estados de ánimo. El acceso y estructura de Twitter es ampliamente utilizado dentro del PLN debido a la gran información subyacente que se puede encontrar o rescatar de cada uno de los usuarios que las conforman y de los diversos enfoques que se le puede dar para resolver problemas de carácter social. Dentro de los lenguajes casados con PLN que permiten manipular información destaca Python [6], el cual cuenta con estructuras de datos eficientes y de alto nivel con un enfoque simple pero efectivo a la programación orientada a objetos.

2.1. Estado del arte

En esta sección, se describe una lista de trabajos relacionados con el presente artículo, que en términos amplios representan aplicaciones PLN. Kulkarni Sagar expone una metodología basada en dos aspectos: 1) la detección de publicaciones maliciosas de forma aislada y sin información previa del usuario y 2) la aplicación de un análisis estadístico de lenguaje para detectar publicaciones maliciosas en el lenguaje “hindi”. Bajo esta metodología se crea un sistema que se puede aplicar a la detección maliciosa en tiempo real, principalmente porque se enfoca en la detección de tweets maliciosos en lugar de cuentas de spam [7].

Henriquez Miranda centra su atención en un sistema de minería de opiniones en idioma español acerca de los comentarios dados por clientes de diferentes hoteles. Estas normas se basan en las evaluaciones hechas en las dimensiones de valencia, excitación y el dominio. Como resultado, se consiguió un sistema más que aceptable, alcanzando una precisión del 94% para el dominio de turismo en hoteles [8].

Por otra parte, Villena-Roman Julio construyó un sistema automático de almacenamiento, análisis y visualización de información semántica extraída de mensajes de Twitter, diseñado para proporcionar a las administraciones públicas una herramienta para analizar los patrones de comportamiento de los ciudadanos, su opinión acerca de los servicios públicos, la percepción de la ciudad, etc. [9].

Las 3 referencias anteriores, a pesar de no pertenecer a PLN para problemas delictivos, han sido de apoyo para las autoras en la comprensión de las herramientas para análisis de los datos y la interpretación de la información expresada en lenguaje natural.

3. Primera aproximación: extracción de información

Para efectos de comparación, se consideró un reciente trabajo que recoge datos de redes sociales para el análisis de los sucesos delictivos que ocurren diariamente donde se hace uso de PLN. En tal trabajo, los autores exponen un análisis sobre información extraída de Twitter para detección de actos delictivos y sensaciones de inseguridad de la población en Barcelona. Sin embargo, en su conclusión, señalan que “Twitter” no funciona satisfactoriamente como herramienta de extracción de información para el problema que abordaron dado que la información tratada no puede ser exacta ni totalmente confiable debido a que el tiempo de estudio fue de solo una semana [10].

Tabla 1. Frecuencia de palabras.

Palabra	Descripción	Repetición
Mano armada	Herramienta de asalto	21
Disparo	Herramienta de asalto	5
Arma blanca	Herramienta de asalto	7
Arma de fuego	Herramienta de asalto	14
Motocicleta/Coche	Herramienta de asalto	2
Ciudad	Lugar de asalto	6 (Córdoba, Hidalgo, Puebla, Cuernavaca, Apizaco, Xonaca)
Tienda comercial	Lugar de asalto	3
Vía pública	Lugar de asalto	3
Transporte público	Lugar de asalto	1
Inocente asesinado	Resultado	16
Inocente herido	Resultado	11
Captura	Resultado	6
Fuga	Resultado	19
Persecución	Resultado	3
Asalto frustrado	Resultado	4
Resistencia	Defensa	2

Ante esta postura, en este artículo para minimizar tal complicación, se han hecho descargas de Twitter durante más de dos meses, lo cual, si bien no garantiza completamente la calidad de los datos, si ayuda a reforzar la fiabilidad de los mismos.

Esta tarea ha permitido contar con información adecuada y en condiciones de ser procesada para definir tanto diccionarios consistentes como satisfacción del análisis de sentimientos. La información conseguida durante 75 días se comparó con los medios de comunicación comunes (periódicos, páginas oficiales del gobierno mexicano, etc.) y se verificó la similitud con los mismos respecto a su semántica.

3.1. Recopilación de datos

Se inicia el desarrollo del trabajo con una inspección dentro de Twitter sobre la inseguridad reportada por diferentes usuarios de habla hispana alrededor del mundo. Como consecuencia, la búsqueda de información se ubica en el análisis delictivo sobre los asaltos. Tal problema es un reto principalmente para países del tercer mundo porque la inseguridad en estos países se refleja principalmente en el robo y/o hurto a personas en diferentes espacios abiertos (vía pública, en el transporte público, comercios, etc.).

3.2. Propuesta de banco de palabras para la búsqueda

A partir de la recopilación de datos, fue posible construir un banco de palabras para la búsqueda y extracción de información que fueran consistentes con el problema que nos ocupa. Este banco contiene conjugaciones del verbo “asaltar” (una de las palabras

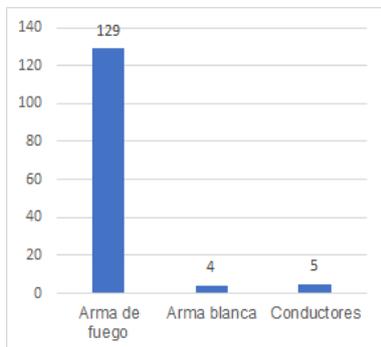


Fig. 1. Gráfica de Herramientas de asalto.

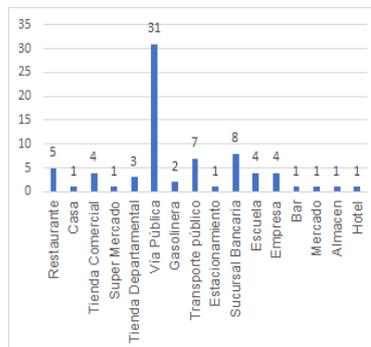


Fig. 2. Gráfica de lugares más asaltados.

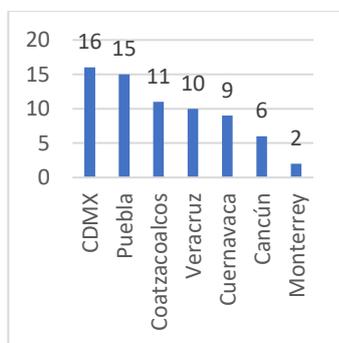


Fig. 3. Gráfica de estados con más asaltos.

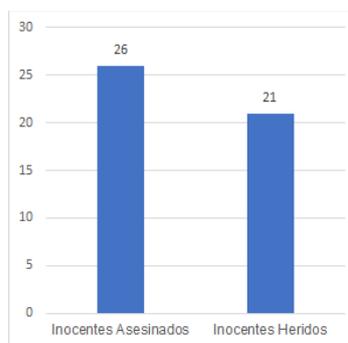


Fig. 4. Gráfica de desencadenamiento de un asalto.

frecuentes), y se suman a estas palabras algunos sinónimos, adjetivos o adverbios relacionados. El vector de palabras para iniciar la extracción formal de información en Twitter se muestra a continuación:

- Palabras: arrestar, arrestaron, asaltantes, asaltar, asaltaron, asalto, atrapar, atraparon, delictivo, delito, inseguridad, ladrones, robar, robaron, robo.
- Hashtags: asaltantes, asaltar, asalto, atrapar, delictivo, delincuentes, delito, inseguridad, ladrones, robar, robaron, robo.

Estas palabras son las candidatas para las búsquedas de los tweets, o bien, ser un hashtag. Para la descarga de los tweets se utilizó la herramienta web Vicinitas, que nos sirve para realizar una descarga de hasta 2080 tweets de los últimos 10 días con claro un parámetro de búsqueda. La información que se obtuvo de la búsqueda fue almacenada en una hoja de cálculo de Excel con los siguientes datos: identificador único del tweet, nombre del usuario, contenido del tweet, fecha de publicación, url, entre otros.

Dado que estas palabras se encuentran en lengua española, los datos recopilados no sólo se concentran en México, sino también, en otros países hispanohablantes. Sin

embargo, como ya se había definido anteriormente, sólo se tomaron en cuenta aquellos Tweets provenientes de la República Mexicana.

Es importante subrayar que, en esta primera aproximación, las descargas de tweets se hicieron de manera periódica durante los meses de enero y febrero de 2020. La información conseguida en estas fechas, permitió el inicio de un análisis del contenido de la descripción de los tweets. Este análisis dio lugar a obtener aquellas palabras que se presentaban de manera repetitiva (sin tomar en cuenta los parámetros de búsqueda). Este vocabulario se consideró para agrupar la información y definir los diccionarios.

De acuerdo a lo obtenido se encontró que se hace mención a: 1) los lugares (espacios públicos o cerrados) en donde ocurren los asaltos/robos, 2) las distintas armas o instrumentos que se usan al llevar a cabo un asalto/robo, así como también el transporte de asalto, 3) la forma de cómo terminan los sucesos después de un asalto/robo y 4) tipo de agresión que reciben las víctimas durante y después del asalto. La frecuencia de las palabras elegidas se muestra en las Figuras 1, 2, 3, 4.

4. Segunda Aproximación: Análisis de Sentimientos

En esta segunda fase se continuó realizando las descargas de los Tweets durante otros dos meses más y se añadieron los registros a los archivos previos. Pero para estas descargas se implementó otro procedimiento distinto al utilizado en las descargas de los Tweets de la primera aproximación. Para ello, se desarrolló un programa desarrollado en Python y usar la API de búsqueda de Twitter, Tweepy. Este programa tiene como ventaja en comparación a Vicinitas que controla con mayor cuidado el filtro de las descargas. La desventaja es el tiempo de uso, por ejemplo, si las descargas rebasan cierto número de peticiones, se deben esperar al menos 15 minutos antes de regresar a obtener información, lo cual puede llegar a generar conflicto para usuarios que requieran descargas masivas de datos. Otro conflicto es la aleatoria negación de la autorización de la cuenta de desarrollador de Twitter.

Una vez que se cuenta con descargas de tweets, se procedió a la de limpieza de datos estableciendo un estándar de la estructura de los tweets. El procedimiento para homogeneizar la información consistió en reunir todos los datos de una palabra en un solo archivo de manera periódica, en plazos de semanas, meses o quincenas. Por ejemplo, si se define un espacio catorcenal de descargas y se encuentra la palabra “asaltar” en ese lapso durante enero, los 14 archivos relacionados se deben conglomerar en uno solo archivo. Después se eliminan los saltos de línea innecesarios de los Tweets o los espacios dobles para ordenarlos de manera ascendente utilizando la fecha. El proceso continúa eliminando los “Re-Tweets” y evitar información repetida (como las tuplas repetidas en el modelo relacional de Bases de Datos).

Este proceso es importante porque nuestro interés es conocer el tipo de asalto, la frecuencia con la que se cometen actos delictivos en vez de contar cuántas veces se compartió un mismo Tweet. Por otro lado, se retiró de los archivos la información adicional que antes fue de utilidad, pero que en este paso ya no es necesaria (Id del Tweet, fecha y hora de este, usuario, etc.), obteniéndose así solo texto del Tweet. De la misma manera se suprimen “emojis” o “caracteres especiales” de cada uno de los Tweets, lo cual fue conseguido desarrollando un programa en C#. Este quehacer fue

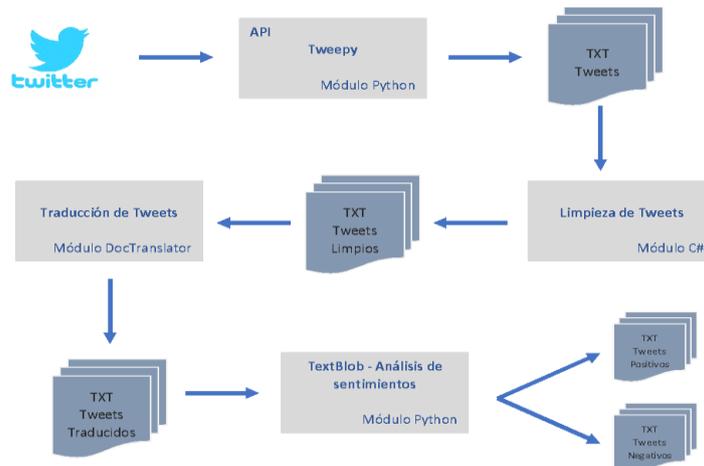


Fig. 5. Proceso de extracción y recolección.

mensual e incluye enero, febrero, marzo y abril de 2020, con 28 palabras de búsqueda en las descargas.

El procedimiento implicó el desarrollo de un programa en Python de utilidad posterior en el análisis de sentimientos. Este programa se encarga de leer en una subrutina los archivos de texto plano (.txt) que se obtuvieron (28 archivos por mes, durante 4 meses). Un proceso posterior toma uno a uno los Tweets y los traduce a inglés, esto dado en consecuencia de que la librería que se utilizó para Analizar Sentimientos (librería TextBlob) necesitaba manejar los Tweets en este lenguaje. Además, es importante enfatizar el hecho de que no se encontraron librerías de PLN que permitieran realizar el análisis de sentimientos para textos en español.

Sin embargo, la traducción de los tweets se interrumpió debido a que la librería utilizada (Google Translate [11]) había alcanzado el número límite de consultas, entonces se optó por traducir manualmente los Tweets en una página en Internet gratuita (onlinedoctranslator.com) todos los archivos de texto (28 archivos por mes, durante 4 meses). Cabe mencionar que se está consciente de que al utilizar herramientas de traducción siempre existe un margen de error en el resultado que proporcionan.

Después del trabajo de obtención, limpieza, concentración de resultados y análisis de sentimientos de la información, se obtuvieron los resultados esperados; algunos irrelevantes y otros sorprendieron a los autores porque no fueron considerados, sin embargo, dieron importancia significativa a la conclusión.

En la figura 1 se presenta de manera general proceso de la extracción y recolección de datos que se llevó en este apartado.

La siguiente subsección se encarga de exponer *Análisis de Sentimientos* y de los resultados obtenidos.

4.1. Búsqueda y análisis de sentimientos

En esta sección se presenta el análisis de sentimientos que surge a partir del primer análisis que arrojó la frecuencia de las palabras dentro del tweets. Un nuevo enfoque

```
#Programa de análisis de sentimientos
Importar librerías para manipular archivos y Análisis de sentimientos;
var listaTweetsInglés := archivo de excel con Tweets (Inglés);
var listaTweetsEspañol := archivo de excel con Tweets (Español);
var resultadoPositivo := crear y abrir nuevo archivo de excel;
var resultadoNegativo := crear y abrir nuevo archivo de excel;
var resultadoNeutral := crear y abrir nuevo archivo de excel;
var contador := 0;
var nuevaListaTweetsEspañol := listaTweetsEspañol;
repetir para cada Tweet en listaTweetsInglés
  var analisis := calculaSentimientos(Tweet)
  var polaridad := polaridad de: analisis
  si (polaridad es igual a 0)
    agregar en resultadoNeutral := nuevaListaTweetsEspañol en
  contador;
  FinSi
  si (polaridad es mayor que 0)
    agregar en resultado Positivo := nuevaListaTweetsEspañol en
  contador;
  FinSi
  si (polaridad es menor que 0)
    agregar en resultadoNegativo := nuevaListaTweetsEspañol en
  contador
  FinSi
  contador := contador + 1
FinRepetir.
```

Algoritmo 1. Código para la detección de sentimientos.

Tabla 2. Ejemplos de tweets por polaridad.

< 0	<p>En violento #asalto balean a pasajero de la Línea Periférico antes de la #RectaCholula, los asaltantes se dieron a la fuga. Pasajeros lo auxiliaron mientras el chofer los traslado a un hospital de San Pedro Cholula. #TransportePublico #Puebla #PueblaNoticias Pol: -0.8</p> <p>La situación se ha vuelto cada vez peor! Aparentemente un pasajero, que viajaba en la parte trasera, se resistió y discutió con uno de los asaltantes que realizo por lo menos un disparo Pol: -0.25</p> <p style="text-align: center;">Neutral</p>
> 0	<p>Tras #Persecucion en #Apizaco, policías #Aseguran a probable responsable de #Asalto a #Transeunte Pol: 0.1</p> <p>¡Adoro los finales felices! "Hombre persigue a #asaltantes que le quitaron 200mp; ellos chocan su moto y mueren." Pol: 0.75</p>

e interpretación ha sido centro de discusión en esta parte: Se señala una reacción al haber sido víctima de algún asalto/robo.

El análisis de sentimientos en este contexto fue desarrollado con Python y para ello es necesario traducir al idioma inglés la información recogida, tal como se describió al principio de esta sección.

TextBlob [12] es una biblioteca dedicada al PLN y es necesaria en el análisis de sentimientos. Esta analiza una cadena de texto generando un peso que se interpreta como polaridad, la cual es la emoción expresada dentro de la oración y tiene un rango de -1.0 a +1.0, lo que significa un resultado interesante y relevante en el artículo.

Tabla 3. Frecuencia de las palabras negativas y positivas.

N°	Negativas		Positivas	
	Palabras	Frecuencia	Palabras	Frecuencia
1	Odio	614	Vive/Vida	254
2	Arma(s)	348	Responsables(s)	230
3	Miserables	342	Justicia	219
4	Mierda	309	Libre	193
5	Corrupción	282	Atrapar	191
6	Violento	267	Cárcel	159
7	Maldito	243	Defender	121
8	Víctima(s)	242	Preso(s)	112
9	Inseguridad	218	Denunciar	111
10	Pinche	167	Arresto(s)	108
11	Terrible	144	Detenido(s)	103
12	Imbécil	142	Libertad	101
13	Peligroso	76	Inocentes	99
14	Dolor	50	Respeto	97
15	Desgracia	44	Feliz	96
16	Infeliz	41	Merecer	94
17	Tristeza	41	Suerte	92
18	Pánico	37	Prevención	91
19	Impotencia	32	Detención	81
20	Coraje	31	Captura	76
21	Ansiedad	20	Asegurar	63
22	Lesiones	20	Protección	42
23	Vándalos	20	Esperanza	41
24	Muerte inocente	20	Vigilancia	22

La polaridad que asigna TextBlob a la información es señalada en tres categorías: Positiva, Negativa y Neutral. Para propósitos de este documento, es válida la etiqueta Positiva y Negativa.

En el Algoritmo 1 se presenta el pseudocódigo implementado para el Análisis de Sentimientos.

Hemos insistido que los Tweets son una cadena de texto que ha sido depurada y expresada de manera adecuada para ser evaluada en el análisis de sentimientos. Cada Tweet presentaba una ponderación diferente.

En la tabla 3 se pueden observar algunos ejemplos representativos. Es oportuno señalar que Tweets neutrales no son relevantes en este estudio, lo cual responde a que, en la creación de los diccionarios, una disyuntiva de positivo y negativo era de utilidad.

De acuerdo con lo anterior se concluye:

Valores positivos:

- Cuando existe un asalto frustrado, la mayoría de las veces son las mismas víctimas las que se logran defender, o estas reciben apoyo de otros civiles.
- Existen casos en los que los asaltantes salen heridos por alguna situación, y las víctimas toman una actitud gratificante ante ello.
- Después de un asalto, es un resultado de alivio si las pérdidas solo fueron cosas materiales.

Valores negativos:

- En la mayoría de los asaltos, las víctimas demuestran su inconformidad con el sistema de vigilancia/seguridad y con los responsables de la supervisión.
- La mayoría de las emociones reflejadas en los Tweets son enojo, tristeza, frustración.
- En casi ninguno de los asaltos, las víctimas presentaron la denuncia ante la fiscalía correspondiente. Cuando las personas asaltadas denunciaron, su emoción es desalentadora porque el proceso fue largo y no satisfactorio.
- Se encontraron algunos casos donde la policía o militares fueron los que agredieron a los civiles.
- En la mayoría de los asaltos, la policía no brinda el apoyo necesario ni actúa de la manera correspondiente. En casos particulares, el apoyo fue completamente nulo.

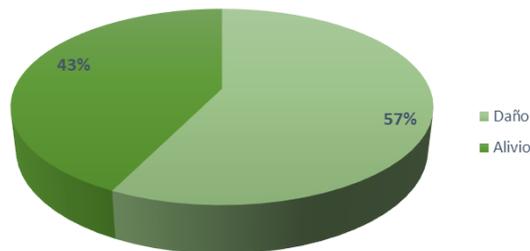


Fig. 6. Gráfica de porcentajes de Daño y Alivio.

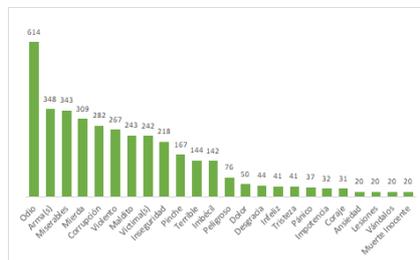


Fig. 7. Gráfica de frecuencias de Daño.

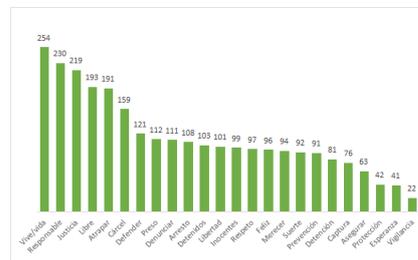


Fig. 8. Gráfica de frecuencias de Alivio.

5. Conclusiones y trabajo a futuro

En este trabajo, una investigación sobre los asaltos que ocurren en la sociedad de la república mexicana por medio de la red social Twitter ha sido desarrollado. Considerando que Twitter es flexible y permite generar datos con gran variabilidad de opiniones, se eligió a esta red social como objeto de estudio extracción de datos.

Una bondad de Twitter es la posibilidad es ajustar con PLN las opiniones de usuarios y establecer tendencias, describir sucesos o noticias identificadas como hashtags. propuesta resulta flexible para estos casos.

La gran cantidad de datos que se encontró dentro de los tweets (128,000 seleccionados de 300,000) fue el resultado del esfuerzo de las autoras que permitió alcanzar el objetivo de este artículo.

La gran variabilidad de Tweets, giraban alrededor de mensajes de alerta que se centraban en avisos acerca de asaltos en algún lugar específico y con ciertas características. Este tipo de Tweets son muy importantes porque produce eco en la red para que otros usuarios se protejan y tomen precauciones de un lugar y hora determinados. Otros avisos interesantes son los tweets que escriben personas que fueron espectadoras de un asalto. El banco de preguntas fue el conductor conveniente para una búsqueda de información consistente y correcta.

Alivio y Daño

El análisis de sentimientos hecho en cada uno de los tweets generó dos diccionarios: positivo y negativo. Una revisión y evaluación sobre los diccionarios fue realizada provocando el planteamiento de las siguientes preguntas:

- ¿Qué es “positivo”?
- ¿Qué es “negativo”?

Es sabido que el tema de los asaltos no es algo que podría definirse exactamente como algo bueno o que pueda siquiera acercarse a ser positivo, sin embargo, se admite que palabras aproximadas a positivo puedan ajustarse a la interpretación de Tweets que calificados como “buenos”.

Se ha nombrado Alivio al diccionario positivo, este nombre responde a la acción de consuelo cuando un asalto no fue exitoso para el asaltante o bien, la víctima no sufrió daños graves físicamente o solo fueron pérdidas materiales o cuando en el asalto frustrado el agresor es lesionado.

Para el diccionario negativo denominado Daño, se encontraron diversos identificadores. Este diccionario significa la gravedad de un asalto y los daños, los cuales son traducidos como lesiones/agresiones que ponen en riesgo la vida. Esta situación genera sentimientos negativos similares al dolor, tristeza, odio, resentimiento, impotencia, etc.

Circulan en el análisis de los Tweets de este trabajo, frases alrededor de incertidumbre, decepción, desaliento y pensamientos de ineptitud hacia las autoridades. Estas palabras se encuentran tácitas en las siguientes oraciones: a) Hay más asaltos vía pública que en cualquier otro lugar, b) En la mayoría de los asaltos, el agresor consigue escapar.

En el mes de marzo, algunos delincuentes tomaron ventaja de la situación sanitaria del país y asaltaron casas haciéndose pasar por trabajadores del sector salud o policías. Además, la actividad delictiva no cesó a pesar del escenario actual.

En marzo, algunas víctimas denuncian que fueron asaltados por el hecho de poseer víveres o productos de primera necesidad, esto debido al COVID-19. Se asume que, si hubiera sido posible desarrollar este trabajo en mayo y junio de 2020, se encontrarán muchas expresiones de agresiones a personas portadoras del COVID-19.

La mayoría de los asaltos descritos se llevaron a cabo dentro de México, más que en cualquier otro país hispanohablante.

El trabajo futuro consiste en mapear los resultados de este trabajo en un Sistema de Información Geográfico para mostrar gráficamente la concentración de zonas vulnerables y las frecuencias de las distintas agresiones delictivas.

Referencias

1. Arauz, R.M.L.: Sistema experto de clasificación y apoyo al diseño arquitectónico de edificios inteligentes. México: UNAM (1998)
2. Schalkoff, R. J.: Artificial intelligence: an engineering approach. Michigan: McGraw-Hill (1990)
3. Barr, A., Feigenbaum, E.A.: The handbook of intelligence artificial. Eyrolles, París (1986)
4. Feldman, R., Sanger, J.: The text mining handbook. Cambridge University Press, <http://www.safaribooksonline.com> (2009)
5. Celaya, J.: La Empresa en la WEB 2.0. Editorial Grupo Planeta (2011)
6. Python Software Foundation, 201 (2008)
7. Kulkarni, S.: Detection of Malicious Microblogs from Online Social Networking Sites using Natural Language Processing (NLP) Techniques. In: 5th International Conference On Innovations in IT and Management. Sinhgad Institute of Management and Computer Application (SIMCA) (2020)
8. Sociedad Española para el Procesamiento del Lenguaje Natural: Minería de opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. <https://www.redalyc.org/pdf/5157/515754423002.pdf> (2020)
9. Villena-Román, J., Luna-Cobos, A., González Cristóbal, J.: Análisis semántico de la opinión de los ciudadanos en redes sociales en la ciudad del futuro. Procesamiento del Lenguaje Natural, 53, pp. 159–162. <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5064/2952> (2014)
10. Puertas-Criado, S.: El Uso de Twitter y Tecnologías SIG como herramienta para la detección de victimización y percepción de inseguridad. Universitat Pompeu Fabra Barcelona (2017)
11. Han, S.: GoogleTrans (Versión 2.4.0) <https://pypi.org/project/googletrans/> (2018)
12. Loria, S.: TextBlob (Versión 0.15.3) <https://pypi.org/project/textblob/> (2019)