

# Introducción

Módulo de desarrollo de retos

*Reto Infotec-CentroGeo*

El uso de redes sociales para realizar denuncias es un fenómeno que se ha visto recientemente. Las personas utilizan estos medios dada la simplicidad para hacerlo y el tipo de alcance deseado. La flexibilidad del medio permite que las denuncias sean comunes y es posible contabilizar las denuncias a lo largo de diferentes espacios geográficos de manera sencilla.

Sin embargo, el lenguaje utilizado en redes sociales puede ser informal y complejo, y es común encontrar diversos mensajes que hacen mención a supuestos delitos pero que en realidad pueden no ser una denuncia, o incluso, lenguaje figurado más que hechos.

Este reto busca construir un modelo eficiente y eficaz para la identificación de posibles denuncias de delitos mediante técnicas de procesamiento de lenguaje natural. Como es bien sabido la clasificación de texto puede tener diversos retos como la ambigüedad, la falta de información contextual, la semántica en sí, por ello poder determinar si de manera explícita se menciona la ocurrencia de un delito es muy complejo.

Para ello, se definió un conjunto de datos de tweets en Español donde se clasifican como:

**1) Se menciona la ocurrencia de un delito**

**0) No se menciona la ocurrencia de un delito**

Como se puede observar se trata de un problema de clasificación binaria, y con un conjunto de datos desbalanceado, donde de manera natural, habrá más observaciones de la clase 0 (no ocurrencia de delito).

Las métricas que se emplearán para medir el rendimiento de los modelos de clasificación son: Precision, Recall y F1-Score en la clase positiva.

La comunidad objetivo del reto son estudiantes que están iniciando en el mundo del procesamiento del lenguaje natural y aprendizaje automático para que puedan explorar sus ideas y analizar los resultados que obtienen con ellas. Una vez construido el modelo, la idea es poder determinar con un conjunto de datos mayor, el porcentaje de delitos que han sido explícitamente mencionados durante un periodo de tiempo en twitter, y así tener, por decirlo así, una estadística experimental medible en corto tiempo y con costo bajo.

Los datos con los que se trabajará es un conjunto de datos obtenido de la red social Twitter en el periodo de tiempo del 2016 hasta marzo del 2023, escogidos de forma aleatoria. Este conjunto de datos consta de 1,352 tweets clasificados como tuits negativos (label 0) y 449 tuits positivos (label 1). El corpus está dividido en un conjunto de entrenamiento [delitos\\_train.json](#) y un conjunto de prueba [delitos\\_test.json](#).

**El formato de estos dos archivos es un json por linea**, el json tiene tres llaves; annotations que continen la lista de etiquetas asignadas por tres personas, id es el identificador del tuit y label que es la clase asignada. El siguiente código muestra un tuit etiquetado.

```
{'annotations': [0, 0, 0], 'id': 1152148249157021698, 'label': 0}
```

## Proceso de etiquetado

El proceso de etiquetado se llevo acabo por un grupo de profesores expertos en diversas áreas del conocimiento, como políticas publicas, procesamiento de lenguaje natural, principalmente.

El procedimiento que se siguió para etiquetar un tuit como positivo o negativo fue el siguiente. Cada uno de los etiquetadores leía cada tuit y lo identificaba como positivo si era posible contestar la pregunta: ¿Cuál es el delito? y además se debía cumplir que el mensaje fuera un hecho, es decir, que el objetivo del mensaje, sea sobre el delito, que no sea una opinión y que no se trate de un accidente.

Con el propósito de clarificar estas reglas presentamos algunos ejemplos.

En los siguientes tres ejemplos, identificamos que se está dando una opinión sobre un delito por ese motivo se marca como negativo (0).

- **Este caso tiene mucho que ver con lo que paso en EEUU, Kenosha —Kyle Rittenhouse, el notorio adolescente blanco armado acusado de asesinar a dos personas e intentar matar a una tercera durante una protesta por la violencia policial en agosto pasado, fue declarado inocente.**

- **Así se las gasta la gente del #FrenteAncho... Los bomberos obligados a manifestarse vs Obrador y los disidentes amenazados... Ayer asesinaron a uno de ellos... @IsmaelBombero señalado en video...**

- **Amenazas a la fiscal Ximena Chong : Hagamos memoria - Estallido Social en Chile**

En los siguientes tuits se infiere, por el contexto, que el delito (sí hay delito, clase 1) es el tráfico de personas y conducción en estado de ebriedad dado que se está usando una patera y se menciona un chofer.

- **Detenido el organizador de un viaje en patera en el que murieron 11 inmigrantes**

- **Chofer de autobús escolar ebria es detenida por estudiantes valientes**

En los siguientes mensajes se identifica el delito, pero el objetivo del mensaje es comunicar otro evento que no es un delito (etiqueta 0).

- **La Policía Nacional homenajea a cuatro agentes asesinados en octubre**

- **HOY SE CUMPLEN 18 AÑOS DEL ATAQUE TERRORISTA A LOS E.U.A. Los atentados del 11 de septiembre de 2001 fueron una serie de cuatro atentados terroristas suicidas cometidos en la mañana del martes #September11 #torresgemelas**

**- Cuerpos de seguridad del estado desmantelaron la banda de roba coches**

**- Radican cargos en ausencia a mujer por tentativa de asesinato en Aibonito**

En los siguientes ejemplos se menciona que las personas fueron atropelladas lo cual se considera como delito, pero en el primero se puede suponer que fue un accidente y en el segundo no hay información para suponer que el evento fue accidental.

**- Accidente deja un muerto en #Galeras #SUCRE Benito Navarro iba en una moto de Sincé hacia Galerías, paró en una orilla de la carretera a orinar, otra moto que iba detrás lo atropelló, lanzándolo hacia la vía, un camión que venía detrás le provocó heridas mortales en la cabeza.**

**- Lamentablemente el conductor de una motocicleta perdió la vida al ser atropellado por un camión, esto en Avenida El Rosario casi con calle Cultura Norte Colonia El Rosario Alcaldía de Azcapotzalco.**

El proceso de etiquetado que se siguió se dividió en tres etapas, en las primeras dos etapas solo se involucró una persona y en la última etapa se involucrarán tres personas.

### **Primera etapa**

La primera etapa consistió en buscar mensajes (tuits en el universo de tuits públicos en el rango de fecha mencionado) que tuvieran las palabras robar, asaltar, quitar, secuestrar y asesinar o q-gramas de estas palabras. Se creó una bolsa de palabras y q-gramas y se seleccionó, con la similitud coseno, los tuits más similares a la consulta creada con las palabras en días seleccionados al azar. Se recuperaban los 8 tuits más similares por día y se etiquetaron como positivos aquellos que mencionan algún delito y negativo aquellos que no tenían mención de delitos, siguiendo las reglas mencionadas anteriormente.

### **Segunda etapa**

Cuando en la primera etapa se tuvieron etiquetados un par de decenas de la clase positiva, se inició con la segunda etapa. En esta etapa se entrenó un clasificador donde el texto era representado en un espacio de emoticones (ver [EvoMSA](#)) y el clasificador era un máquina de soporte vectorial lineal. Al azar se seleccionaban todos los tuits recolectados en un hora particular, de esos tuits se seleccionaban 2048 tuits, etiquetando los 8 tuits con el máximo valor de la función de decisión del clasificador. Este proceso se continuaba hasta que se terminaban todos los tuits recolectados en la hora, o hasta que se repetía algún tuit ya etiquetado. Finalmente, el modelo se actualizaban de manera automática cuando se seleccionaba otra hora. El proceso mencionado anteriormente se realizó hasta tener etiquetados 500 tuits con la clase positiva y se obtuvieron 1,500 tuits de la clase negativa.

### **Tercera etapa**

En la tercera etapa se etiquetó el conjunto de 1,800 tuits por tres personas siguiendo las reglas mencionadas anteriormente. Este será el conjunto etiquetado para éste reto.

**Sitio de delitos**

**<https://ingeotec.github.io/Delitos/>**