PUEDJS
PROGRAMA UNIVERSITARIO DE ESTUDIOS SOBRE DEMOCRACIA, JUSTICIA Y SOCIEDAD

Tlatelolco LAB

DOCUMENTO DE TRABAJO 11

Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad | FEBRERO 2023

Estrategias metodológicas

Análisis de Sentimiento en Twitter

Incorporación de hashtags y corpus locales de entrenamiento.

Martín Zumaya
Luis Ángel Escobar
Diego Espitia
Autores

FEBRERO 2023

Documento de trabajo 11

Análisis de Sentimiento en Twitter Incorporación de hashtags y corpus locales de entrenamiento.

Este documento de trabajo fue elaborado por investigadores del Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad (PUEDJS) de la Universidad Nacional Autónoma de México.

Primera edición, Febrero 2023

Programa Universitario de Estudios Sobre Democracia,
Justicia y Sociedad, Torre UNAM-Tlatelolco, Piso 13
Ricardo Flores Magón número 1,
Colonia Nonoalco Tlatelolco Alcaldía Cuauhtémoc,
Código Postal 06995, Ciudad de México
www.puedjs.unam.mx

Cómo citar:

Zumaya Martín, Luis Ángel Escobar, Diego Espitia (2023), "Análisis de Sentimiento en Twitter Incorporación de hashtags y corpus locales de entrenamiento.", Documento de Trabajo núm. 11. PUEDJS, UNAM, México, 26 páginas.

El presente documento es producto de una investigación realizada en el marco de los Programas Nacionales Estratégicos del Consejo Nacional de Ciencia y Tecnología (PRONACES-CONACYT). Agradecemos al CONACYT por el generoso apoyo brindado en 2022. El contenido y las opiniones son responsabilidad exclusiva de los autores.
| **contacto: contacto.puedjs@gmail.com**

ÍNDICE

Presentación

Twitter se ha convertido en una de las plataformas digitales de discusión e información más populares en donde se reacciona y comenta de manera casi inmediata a diversos acontecimientos sociales y políticos. Identificar y analizar el sentimiento de las publicaciones en la plataforma alrededor de estos eventos es útil para caracterizar la percepción, tanto de manera positiva como negativa, de quienes participan en la discusión. Sin embargo, dada la naturaleza digital y la gran cantidad de información generada en las plataformas, es necesario utilizar herramientas computacionales y estadísticas para su análisis, una de ellas es el Análisis de Sentimiento (AS) por medio de modelos y herramientas del Aprendizaje de Máquina.

Previamente en el documento de trabajo: "Análisis de Sentimiento en Twitter. Aproximaciones con métodos de Aprendizaje de Máquinas" [3], realizamos una primera aproximación al análisis de sentimiento aplicado a textos generados en redes socio-digitales, específicamente en la plataforma Twitter, en donde pudimos estimar el sentimiento general de un conjunto de tweets mediante la implementación de un par de modelos: uno de clasificación compuesto por una Red Neuronal Artificial (RNA), y otro basado en léxico. Así mismo, partiendo de metodologías que desarrollamos para el análisis de comunidades [4], junto con estos modelos de aprendizaje máquina, evaluamos la polarización de comunidades de cuentas en Twitter por medio del AS de los textos de los tweets asociados a ellas y la diferencia entre el número de textos clasificados como positivos y negativos.

Si bien ésta primera aproximación al AS ha sido satisfactoria, en el sentido de que los modelos implementados nos permiten caracterizar de manera general el sentimiento de un conjunto de publicaciones, aún quedan detalles a considerar que pueden ser útiles para realizar un análisis más preciso. Por ejemplo, dado que el modelo de RNA previo fue entrenado con un corpus anotado de textos de español de España, consideramos que el uso de un corpus de español Latinoamericano puede capturar de mejor manera las expresiones locales y por lo tanto, permitirnos realizar una mejor aproximación del sentimiento de los tweets, siendo uno de los objetivos principales de este documento evaluar el desempeño y efecto de modelos de AS entrenados con corpus con expresiones locales.

Para explorar esta situación, en este documento implementamos modelos prototipo que consideran no solo el español de España sino también el de México y Colombia, en los que observamos que los modismos y palabras que dependen del contexto socio-político de cada país tienen un impacto en la evaluación de los textos, ya que un conjunto de publicaciones clasificadas como positivas por el modelo de RNA entrenado con textos de español de España, pasan a ser clasificados como negativos. Lo cual se debe a que los textos incluyen palabras con connotación negativa de uso local y que no están presentes en los ejemplos del corpus de español de España. Como mostramos también en este documento, ésta situación, nos permite construir léxico de mediante la identificación de las palabras relevantes con un análisis de TF-IDF de los ejemplos que cambiaron de clasificación.

En este documento identificamos que los hashtags en los tweets del corpus que utilizamos para entrenar nuestros modelos de clasificación, no contribuyen al AS, por lo que otro de los objetivos de este trabajo es el de explorar maneras de incluirlos en el proceso de clasificación evaluándolos de manera independiente al texto de los tweets, con la hipótesis de que el contraste entre la evaluación del texto y los hashtags, puede ser una vía para identificar el uso irónico de los hashtags en las publicaciones que se comparten en Twitter.

En suma, en este documento mostramos que el corpus de entrenamiento de los modelos de AS tiene un efecto importante en la clasificación que realizan los modelos, y que es importante incorpar textos anotados con expresiones locales para tener una mejor estimación del sentimiento de un conjunto de publicaciones. Así mismo, exploramos estrategias para incorporar los hashtags (o etiquetas) que acompañan los textos, para complementar la información utilizada por los modelos de clasificación.

La implementación y exploración de estas líneas de investigación, tanto en este documento como en el futuro, nos permiten seguir estudiando el uso del lenguaje y la percepción de distintos acontecimientos sociales y políticos en las redes socio-digitales.

1. Introducción

Las redes socio-digitales se han convertido en uno de los espacios de discusión más importantes y populares, en los que el acceso inmediato a la información hace posible reaccionar a eventos sociales y políticos apenas estos suceden, e interactuar con la información y comentarios que se generan alrededor de ellos

En este sentido, nos enfocamos al análisis de eventos focales [4], es decir, a aquellos eventos de relevancia social y política alrededor de los que se observa una mayor actividad y reacciones en las plataformas socio-digitales, y que pueden tener tanto un impacto en procesos democráticos como en la generación de sentidos comunes y la percepción de la ciudadanía.

Analizar las millones de publicaciones que reslizan las personas usuarias en las plataformas de las redes socio-digitales sobre diversos temas es una tarea imposible de realizar de manera manual. Por lo que es necesario utilizar herramientas computacionales y estadísticas como el AS mediante modelos de Aprendizaje de Máquina (o Machine Learning), pues nos permite procesar y analizar de manera eficiente grandes cantidades de datos.

En nuestro caso específico nos enfocamos en el AS de textos de publicaciones en la plataforma Twitter (o tweets) y al contenido generado por comunidades de cuentas identificadas por sus patrones de comportamiento e interacción [4]. El énfasis en analizar ésta plataforma se debe a que se ha convertido en un espacio en donde se discuten temas sociales y políticos actuales y por que permite la recopilación y análisis sistemático de la información y contenido que se genera en ella mediante el desarrollo y aplicación de herramientas computacionales y de análisis estadísticos.

En este documento trataremos cuatro eventos focales ocurridos en 2022 listados en la Tabla 1.1, los cuales consideramos relevantes por las siguientes razones: las elecciones en Colombia nos permiten observar diferencias con el lenguaje utilizado en México alrededor de un proceso electoral, lo cual nos permite evaluar el desempeño de nuestros modelos de clasificación en función del corpus de entrenamiento; la discusión alrededor de la consulta ciudadana de Revocación de Mandato tuvo presencia en Twitter por un periodo largo de tiempo, lo que nos permite obtener un conjunto de datos amplio para el entrenamiento y evaluación de los modelos que implementamos en este documento; por otro lado, el incidente en el estadio corregidora es un evento social con menor contenido político, por lo que tanto las comunidades de cuentas alrededor de éste, como el lenguaje utilizado es distinto respecto a otros temas y eventos que hemos analizado previamente; por último las elecciones en México destacan por la presencia de palabras y hashtags que hacen referencia al fenómeno del narcotráfico, un tema arraigado en el México actual.

El modelo de AS compuesto por una RNA implementado anteriormente en el documento de trabajo “Análisis de Sentimiento en Twitter. Aproximaciones con métodos de Aprendizaje de Máquinas” [3], fue entrenado con un corpus creado en 2012 para el Taller de Análisis de Sentimientos (TASS) de la Sociedad Española de Procesamiento de Lenguaje

Evento	Fechas
Elecciones presidenciales de Colombia	29 de mayo al 19 de junio del 2022
Consulta ciudadana de revocación de mandato	10 de abril del 2022
Incidente de violencia en el estadio Corregidora en Querétaro	5 de marzo del 2022
Elecciones intermedias en México	5 de junio del 2022

Tabla 1.1: Eventos focales analizados para la elaboración de este documento y fecha de inicio de la extracción de datos.

Natural (SEPLN) [1], el cual cuenta con más de 60,000 mil tweets evaluados como positivos, negativos o neutros. El desempeño de dicho modelo ha sido satisfactorio pues es capaz de detectar correctamente el sentimiento general de un conjunto de tweets, sin embargo, consideramos que la aplicación de este modelo, entrenado con tweets en español de España, a eventos sucedidos en México puede ser limitada ya que dicho corpus de entrenamiento no contiene expresiones y particularidades locales del lenguaje, como el uso de palabras con connotación negativa en el español mexicano que no corresponden al uso que se les da en el español de España.

Por tanto, uno de los objetivos de este documento es el de evaluar el desempeño y efecto de modelos de AS entrenados con distintos corpus que incluyen expresiones locales, mediante la comparación de la clasificación realizada por estos modelos con el modelo previo entrenado con el corpus general TASS.

Tanto el modelo implementado previamente como los que consideramos en este documento, asignan a cada texto perteneciente a un tweet, t_i, la probabilidad de que pertenezca a la clase positiva y negativa:

$$M(t_i) = (p_i^+, p_i^-) \tag{1.1}$$

De tal manera que para asignar una categoría a cada texto, $c(t_i)$, consideramos la probabilidad mayor asignada por el modelo, es decir, el texto de un tweet es clasificado como **negativo** cuando $p_i^- > p_i^+$ y **positivo** cuando $p_i^+ > p_i^-$, lo cual puede expresarse mendiante la diferencia entre las probabilidades de pertenencia $\delta_i^\pm = p_i^+ - p_i^-$ como:

$$c(t_i) = \begin{cases} \text{Positivo}, & \text{si } \delta_i^\pm > 0 \\ \text{Negativo}, & \text{si } \delta_i^\pm < 0 \end{cases} \tag{1.2}$$

De tal forma que, aunque los corpus de entrenamiento para los modelos que implementamos sean distintos, el resultado de su aplicación a los datos será la misma, es decir, asignan un par de probabilidades de pertenencia para cada texto evaluado que utilizamos para asignarles una categoría.

En resumen, en este documento exploramos dos aspectos del AS de textos de publicaciones en Twitter: el efecto del entrenamiento de un modelo de AS con un corpus que incluye expresiones locales, y una estrategia para la incorporación de hashtags en el AS, las cuales desarrollamos en las siguientes secciones.

2. Incorporación de expresiones locales en el corpus de entrenamiento.

Las diferencias culturales que existen fuera de las redes y plataformas socio-digitales y que se expresan en el lenguaje popular generan, incluso dentro de un mismo idioma, modismos que tienen impacto en la emocionalidad y el sentimiento de nuestras expresiones. En una conversación cara a cara, la emocionalidad y el sentimiento de una expresión puede identificarse haciendo uso de elementos adicionales como el tono de la voz o el lenguaje corporal, de manera que es posible que personas de diferentes culturas puedan relacionar los modismos con la emoción que el hablante busca transmitir aunque no los conozcan. Sin embargo, esta situación es más complicada cuando analizamos solamente el texto mediante herramientas computacionales sin acceso a un contexto adicional, ya que para aproximar el sentimiento de los textos y los diferentes modismos en ellos, es necesario contar con un corpus previamente evaluado que contenga este tipo de expresiones, para después poder utilizarlo en el entrenamiento de un modelo de AS.

El corpus utilizado para entrenar nuestro modelo es el corpus creado para el taller del 2012 de la Sociedad Española de Procesamiento del Lenguaje Natural, al que de ahora en adelante nos referiremos como el corpus general, que si bien ha sido útil, el hecho de que los tweets del corpus estén escritos en un español que ignora muchos de los modismos de uso común en México y además de hace más de 10 años puede generar un sesgo en la clasificación de textos más recientes realizada por el modelo.

En este documento consideramos dos extensiones del corpus general TASS de 2012 con corpus adicionales que incluyen tweets con español mexicano y colombiano, esto con el objetivo de evaluar el efecto de la incorporación de expresiones locales en el entrenamiento de un modelo de AS.

2.1. Extensión del corpus de entrenamiento con ejemplos del español mexicano

Con el objetivo de que nuestro modelo de AS sea capaz de reconocer modismos del español mexicano, extendemos el corpus general TASS con el que entrenamos el modelo original, el cual cuenta con 60,000 tweets, con un corpus adicional también generado para el TASS de 2019 que incluye 1,500 tweets evaluados con contexto mexicano. De tal manera que podemos generar dos modelos de AS entrenados con dos corpus distintos:

- Un modelo de AS entrenado con el corpus general TASS.
- Un modelo de AS entrenado con la combinación del corpus general y el que incluye

ejemplos locales, para obtener uno nuevo que considere algunas expresiones propias del español mexicano, el cual denominamos como TASS_MX.

Una vez que se ha realizado el entrenamiento y optimización de estos modelos, siguiendo el mismo procedimiento descrito en [3], evaluamos su desempeño al comparar los resultados obtenidos con cada uno de ellos.

Para el caso del modelo entrenado con ejemplos de español mexicano, consideramos como eventos focales: los actos de violencia en el estadio Corregidora de Querétaro ocurridos el 4 de abril de 2022, alrededor del cual recopilamos 480,762 tweets publicados entre el 4 de abril y el 9 de abril del 2022 que utilizaron los hashtags #SiHayMuertos, #VerguenzaNacional, #17Muertos, #sihubomuertos, entre otros, o las palabras clave Querétaro y Corregidora, y el ejercicio de la consulta ciudadana de Revocación de Mandato, llevada a cabo el 10 de abril de 2022, alrededor del cual recopilamos 5,191,502 tweets publicados entre el primero de enero y 27 de abril del 2022 que utilizan hashtags como #LigadeGuerreros, #RevocaciondeMandato, #RevocacióndeMandato, #YoDefiendoalINE, #QueSigaAMLO.

Para evaluar el efecto de los distintos corpus, evaluamos los textos de los tweets recopilados con ambos modelos y comparamos la categoría asignada a cada elemento, lo cual mostramos en la Figura 2.1 [1]. Cada punto en la figura representa un texto evaluado por los modelos, en el eje horizontal se muestra p^- asignada por el modelo entrenado con el corpus general, mientras que en el eje vertical se muestra p^- asignada por el modelo entrenado con el corpus extendido, de tal forma que los textos para los que hubo un cambio en su evaluación se ubican fuera de la diagonal. Como se muestra en la figura, en el caso de los incidentes en el estadio Corregidora en Querétaro observamos que 47,563 tweets tuvieron un cambio de más de 0.5 en la probabilidad de pertenencia p^- asignada por el modelo entrenado con el corpus extendido, es decir, el sentimiento asignado por el nuevo modelo es el contrario al asignado previamente, mientras que para la consulta de Revocación de mandato observamos que 603,391 tweets pasaron de ser clasificados como positivos a negativos. Aunque en ambos eventos existen casos en que cambiaron de ambas categorías, observamos que hay un mayor número de elementos que pasaron de ser clasificados como positivos a negativos. Este cambio de clasificación, como describimos a continuación, en este caso está asociado a la presencia de ciertas palabras en los textos.

En la Tabla 2.1 mostramos algunos de los textos relacionados con el incidente en Querétaro que tuvieron un mayor incremento en p^- y que pasaron de ser positivos a negativos con la evaluación del modelo entrenado con el corpus extendido. Observamos que los textos contienen la palabra **pinche**, una palabra que no está presente en los elementos del corpus general compuesto por expresiones del español de España, en el que solamente aparece una vez con un significado diferente al que se utiliza de manera coloquial en México: ***"Please, que nadie pinche el enlace que un tío muy listo está enviando como spam desde mi Twitter…y gracias al tío listo!!"***, en este caso hace referencia a pinchar o pulsar, a diferencia del corpus de tweets mexicanos en el que aparece 17 veces, todas evaluadas de manera negativa.

[1] *La figura comparando la probabilidad de pertenencia a la clase positiva es simétrica, por lo que solamente mostramos la figura para la categoría negativa.*

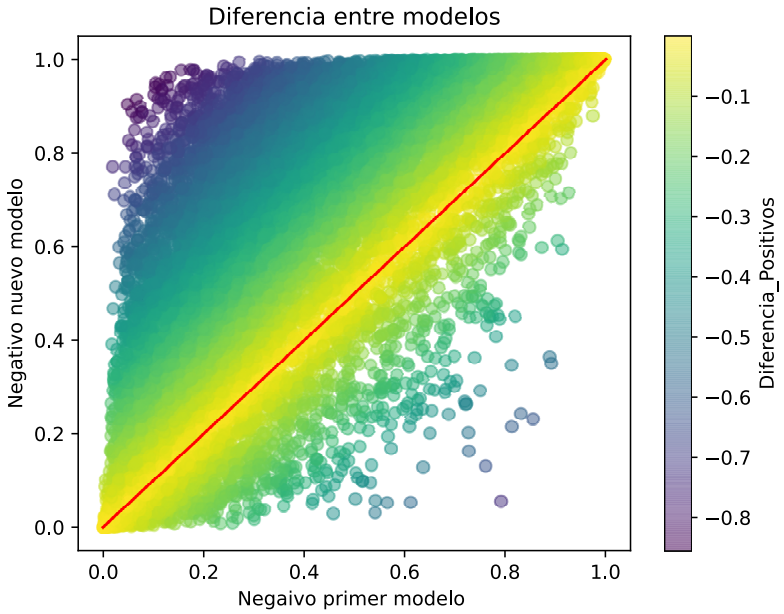

Figura 2.1: Comparación de la evaluación de sentimiento de los textos de los tweets asociados a los incidentes en el estadio Corregidora. El eje horizontal corresponde al modelo entrenado con el corpus general y el eje vertical al modelo entrenado con el corpus extendido, cada punto representa un tweet y el color representa la diferencia entre las probabilidades de pertenencia asignadas por los modelos.

Texto	Positivo Original	Negativo Original	Positivo Nuevo	Negativo Nuevo
Ya te corregí el tweet, pinche Epigmenio. https://t.co/oLUOBjKUmL	0.957412	0.0489678	0.0860086	0.903993
Pinche afición de rancho!!! #Corregidora https://t.co/iRMvp1KMjD	0.942738	0.0659092	0.0816018	0.914504
Wtfff con todo lo de #17muertos pinche sociedad dlv	0.94355	0.0576307	0.103698	0.888534
Dice @MikelArriolaP. Mañana domingo me traslado a Queretaro. Primero duermo y me voy desayunado. Pinche cinico #VerguenzaNacional #Queretaro #Atlas	0.886336	0.101872	0.0865695	0.924644

Tabla 2.1: Ejemplos de tweets con mayor cambio en la clasificación de los 2 modelos para el evento de los incidentes en el estadio Corregidora. Observamos como en todos estos ejemplos se utiliza la palabra "pinche" con una connotación negativa.

Incidentes en el estadio Corregidora	Ejercicio de revocación de mandato
asesinos	fracaso
lamentable	extraño
renuncia	pinche
pendejo	feo
pinche	pendejo
asesino	alv
terror	ahogado
tristeza	patadas
muertos	destruyendo
masacre	balazo

Tabla 2.2: Palabras relevantes en el cambio de clasificación de los textos de positivo a negativo identificadas mediante TF-IDF, resultado de la extensión del corpus original con un corpus de textos evaluados con contexto mexicano, para los casos de los incidentes en el estadio Corregidora en Querétaro y de la consulta ciudadana de Revocación de Mandato. Solamente se muestran los verbos, adjetivos y sustantivos que no están relacionados directamente con el contexto del evento.

Para explorar con más detalle las características de los textos que pasaron de ser clasificados como positivos a negativos por los modelos para ambos conjuntos de datos, identificamos las palabras más relevantes en ellos, mediante un análisis de Term Frequency - Inverse Document Frequency, o TF-IDF [2] por sus siglas en inglés.

En la Tabla 2.2 mostramos ejemplos de las palabras más relevantes para el cambio de clasificación de positivo a negativo, podemos observar que éstas incluyen insultos de uso coloquial en México. Cabe hacer notar que en éstas listas de palabras no hemos incluido las que hacen referencia al contexto particular de cada evento (p.e. futbol, corregidora, revocación, mandato, consulta, AMLO, entre otros), ya que dichas palabras, aunque relevantes para la descripción del evento focal particular que se esté estudiando, no expresan un sentimiento definido.

➔2.2. Extensión del corpus de entrenamiento con ejemplos de español colombiano.

De manera análoga al caso del español mexicano, realizamos el ejercicio ahora con español colombiano, es decir, entrenamos un modelo de AS con un corpus extendido compuesto por el corpus TASS general y un conjunto de tweets evaluados con expresiones del español colombiano, el cual denominamos como TASS_COL. Este corpus, disponible en [2], consta de una lista de los números identificadores de 1,030 tweets [3] , de la cual

[2] *TF-IDF es una métrica estadística diseñada para reflejar la importancia de palabras en un conjunto de documentos (o corpus). Los detalles de la definición de ésta métrica se pueden consultar en [2].*

[3] *Todos los tweets (o publicaciones) cuentan con un número identificador único, generado en el momento de su publicación.*

solamente pudimos recuperar 700 al ser rehidratada [4] mediante el acceso académico a la API de Twitter el 16 de julio del 2022 [5]. En este caso consideramos como evento focal las elecciones presidenciales de Colombia que se llevaron acabo en 2 vueltas, la primera el 29 de mayo y la segunda el 19 de junio de 2022, alrededor de las cuales recopilamos 8,385,810 tweets que utilizaron principalmente los hashtags #EleccionesColombia, #PetroPresidente, #FicoPresidente, #PactoHistorico, entre otros.

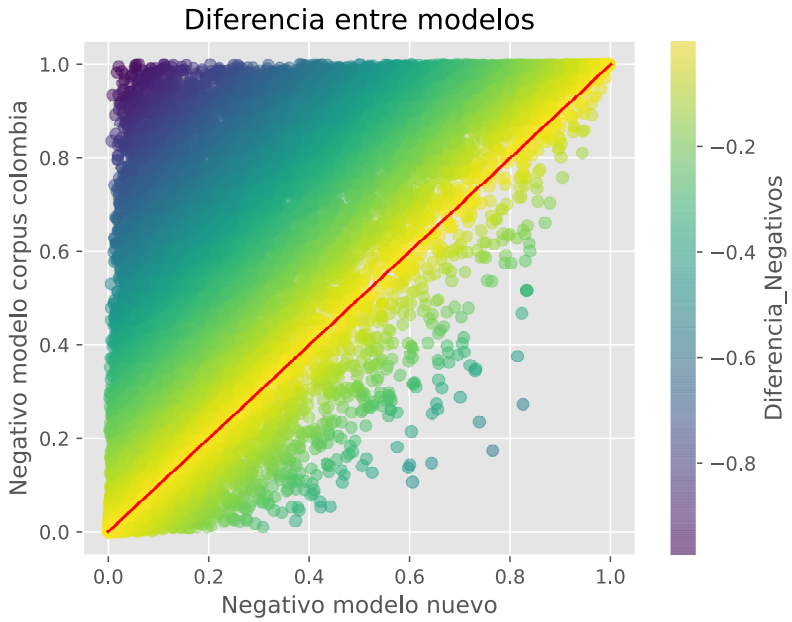

Figura 2.2: Comparación de la evaluación de sentimiento de los textos de los tweets asociados a las elecciones presidenciales en Colombia. El eje horizontal corresponde al modelo entrenado con el corpus general y el eje vertical al modelo entrenado con el corpus extendido, cada punto representa un tweet y el color representa la diferencia entre las probabilidades de pertenencia asignadas por los modelos.

Como se muestra en la Figura 2.2, identificamos 244,203 tweets que pasaron de ser clasificados como positivos por el modelo entrenado con el corpus general, a negativos al ser evaluados con el modelo entrenado con el corpus extendido, algunos de los cuales mostramos en la Tabla 2.3.

Inspeccionando los ejemplos a simple vista, pareciera que la palabra "guerrillero" es una de las que fueron incorporadas en el corpus extendido y que, al estar presente en textos calificados como negativos, contribuyeron al cambio en la clasificación de los textos. Esto se confirma al identificar las palabras más relevantes para el cambio de clasificación, mediante un análisis de TF-IDF de los tweets, las cuales mostramos en la Tabla 2.4, en donde observamos que, a diferencia del caso mexicano, no encontramos palabras de uso

[4] *"Rehidratar" se refiere a la búsqueda y recolección de un tweet o un conjunto de ellos, haciendo uso del acceso a la API de Twitter, por medio de su número identificador único.*

[5] *Existen varias razones por las cuales no es posible recuperar tweets de la base de datos de Twitter: la publicación ha sido eliminada por su autor, la cuenta ha configurado sus publicaciones como privadas o la cuenta que las realizó ha sido suspendida por la plataforma.*

Texto	Positivo Original	Negativo Original	Positivo Nuevo	Negativo Nuevo
Ya te corregí el tweet, pinche Epigmenio. https://t.co/oLUOBjKUmL	0.957412	0.0489678	0.0860086	0.903993
Pinche afición de rancho!!! #Corregidora https://t.co/iRMvp1KMjD	0.942738	0.0659092	0.0816018	0.914504
Wtfff con todo lo de #17muertos pinche sociedad dlv	0.94355	0.0576307	0.103698	0.888534
Dice @MikelArriolaP. Mañana domingo me traslado a Queretaro. Primero duermo y me voy desayunado. Pinche cinico #VerguenzaNacional #Queretaro #Atlas	0.886336	0.101872	0.0865695	0.924644

Tabla 2.3: Tweets con mayor diferencia en la evaluación de los 2 modelos, el del corpus TASS original y el el que combina este con el corpus de tweets en español de México

Elecciones presidenciales en Colombia
guerrillero
jodemos
asesino
imparable
derrota
derrotar
equivoquemos
criminal
agitador
destruir

Tabla 2.4: Palabras relevantes en el cambio de clasificación de los textos de positivo a negativo identificadas mediante TF-IDF, resultado de la extensión del corpus original con un corpus de textos evaluados con contexto colombiano, para el caso de las elecciones presidenciales en Colombia. Solamente se muestran los verbos, adjetivos y sustantivos que no están relacionados directamente con el contexto del evento.

coloquial en Colombia. Sin embargo, hay que tomar en cuenta que el conjunto de tweets con expresiones con español colombiano incorporados al corpus de entrenamiento del modelo, son menos de la mitad que en el caso mexicano y que puede no incluir expresiones más características del lenguaje local.

➔2.3. Comparación entre modelos entrenados con distintos corpus.

Hasta el momento hemos evaluado los modelos entrenados con los corpus extendidos para aplicarlos a conjuntos de datos que corresponden a la misma región, es decir, el modelo entrenado con el corpus extendido con ejemplos de español mexicano lo hemos aplicado a conjuntos de datos asociados a eventos ocurridos en México y el modelo entrenado con el corpus extendido con ejemplos de español colombiano lo hemos aplicado a tweets relacionados con las elecciones presidenciales en Colombia. En esta sección exploramos las combinaciones restantes, para intentar identificar si la diferencia en la evaluación de

los textos depende de su origen. También entrenamos un modelo de AS combinando ambos corpus extendidos en uno solo junto con el corpus general, el cual denominamos TASS_MX_COL, y comparamos los resultados de la aplicación de los modelos a ambos conjuntos de datos.

Específicamente, dado que hemos observado que el efecto principal de la incorporación de textos evaluados con expresiones locales en el corpus de entrenamiento, es el de cambiar la clasificación de los textos de positivo a negativo, calculamos el porcentaje de cambio en el número de textos clasificados como negativos, con respecto al modelo original entrenado con el corpus general.

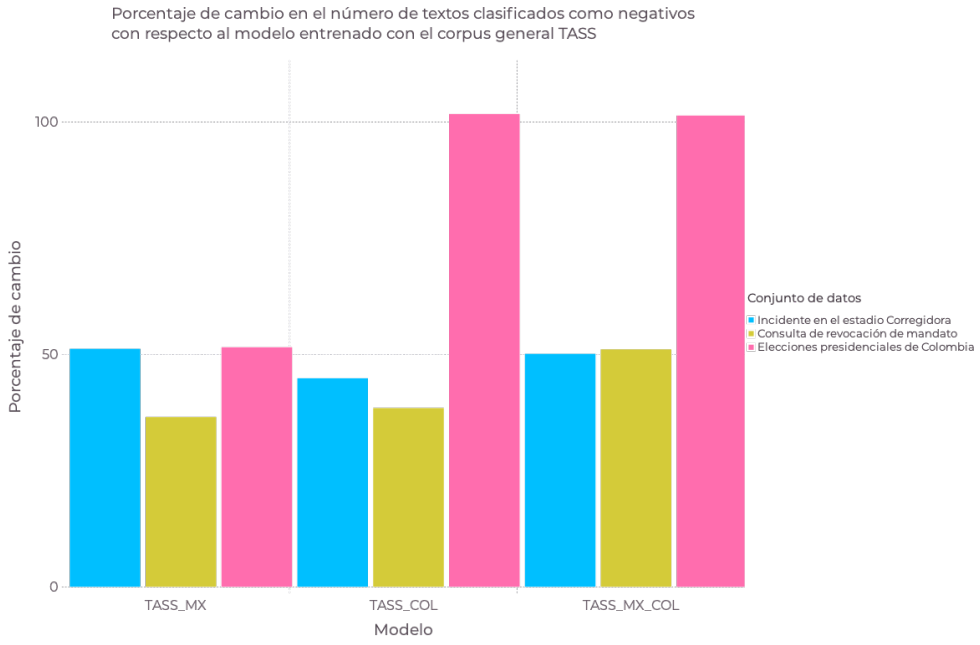

Figura 2.3: Porcentaje de cambio en el número de textos clasificados como negativos, con respecto al modelo entrenado con el corpus general.

Como se puede observar en la Figura 2.3, la inclusión de textos evaluados con contexto latinoamericano, ya sean de México o de Colombia, tiene un efecto en la clasificación de los textos en comparación con el modelo original entrenado con el corpus general TASS. En particular observamos que el corpus extendido TASS_COL, tiene el mayor efecto en el conjunto de datos alrededor de las elecciones presidenciales de Colombia, en donde se tiene un incremento de cerca del 100 % en el número de textos clasificados como negativos, así mismo, observamos incrementos de entre el 30 y el 50 % para el resto de los corpus de entrenamiento y conjuntos de datos.

Estos resultados nos muestran como la ausencia de ejemplos con expresiones locales en el corpus de entrenamiento de los modelos de AS, ya sean positivas o negativas, pueden tener un impacto tanto en la clasificación realizada por el modelo, como en la interpretación de los mismos mediante la sobreestimación de alguna de las dos categorías. En este caso en particular, observamos una subrepresentación de la clase negativa en la clasificación

realizada por el modelo entrenado con el corpus TASS general, en todos los textos asociados a los eventos focales considerados. Así mismo, estos resultados también sugieren que la inclusión en general de expresiones de español latinoamericano (de México y Colombia en el caso de este documento) han tenido un impacto en el desempeño del modelo de AS, ya que como mostramos en la Figura 2.3, tanto la evaluación de textos de México con el modelo entrenado con corpus TASS_COL como la evaluación de textos de Colombia con el modelo entrenado con el corpus TASS_MX, provocan un cambio en la clasificación de un conjunto de textos, lo cual puede deberse a similitudes en las expresiones utilizadas por las personas usuarias de las redes socio-digitales en la región.

2.4. Construcción de un léxico piloto

En esta sección exploramos la posibilidad de construir un léxico compuesto de las palabras representativas utilizadas en los textos de los tweets que tuvieron un mayor cambio en p^- al ser evaluados con el modelo entrenado con el corpus TASS_MX [6]. La hipótesis es que estos textos contienen expresiones más cercanas a los elementos en el corpus extendido que generaron el cambio de clasificación del texto, y por lo tanto representan de mejor manera las expresiones locales.

Por ejemplo, podemos observar que en la Tabla 2.2 hay palabras repetidas en ambos eventos, específicamente: "pinche" y "pendejo". Ambas palabras de uso cotidiano en México y que son utilizadas como insultos o con una connotación negativa, y que no están presentes en el corpus TASS general.

En primer lugar, para identificar a los textos que utilizaremos para construir el léxico de cada uno de los eventos focales que analizamos, consideramos la diferencia entre las probabilidades de pertenencia a la clase negativa asignada por el modelo original y el modelo extendido a cada texto en los tweets de los eventos focales, $\delta_p(t_i)$, definida por:

$$\delta_p(t_i) = p_E^-(t_i) - p_O^-(t_i), \tag{2.1}$$

donde t_i es el i-ésimo texto en el conjunto de datos y $p_O^-(t_i)$ es la probabilidad de pertenencia asignada por el modelo de AS entrenado con el corpus general, y $p_E^-(t_i)$ es la probabilidad asignada por el modelo entrenado con el corpus extendido. Una vez que hemos calculado $\delta_p(t_i)$ para todos los textos, seleccionamos solamente aquellos para los que $\delta_p(t_i) \geq \theta_p$, donde θ_p es un umbral arbitrario, el cual seleccionamos como $\theta_p = 0.8$.

Una vez que hemos obtenido este corpus de textos para cada uno de los eventos focales analizados, lo que podemos expresar como el corpus[7]:

$$\Omega_{\delta_p}(F_i) = \{t_i | \delta_p(t_i) \geq \theta_p\}, \tag{2.2}$$

[6] *Solamente consideramos el caso del corpus TASS_MX y los eventos asociados a México, ya que en el caso de Colombia solamente contamos con datos de un evento focal*

[7] *No confundir con el corpus de entrenamiento del modelo de AS, este corpus se compone de los textos que tuvieron un mayor cambio en la evaluación con el modelo extendido*

donde F_i es el i-ésimo evento focal considerado, procedemos a identificar las palabras más relevantes en cada corpus mediante un análisis TF-IDF. Dado que los eventos focales que consideramos en este documento tienen contextos distintos (procesos electorales, consulta ciudadana y un evento de violencia e indignación social) y que buscamos un léxico general que contenga palabras utilizadas en las Twitter con connotación negativa, construimos el léxico $\mathcal{L}_\Omega$, con la intersección de las palabras relevantes identificadas mediante TF-IDF de todos los eventos, es decir

$$\mathcal{L}_\Omega = \cap_i \text{TF} - \text{IDF}[\Omega_{\delta_p}(F_i)], \tag{2.3}$$

de tal forma que $\mathcal{L}_\Omega$ se compone de las palabras comunes a todos los conjuntos de palabras relevantes, y que pueden tener una connotación negativa.

La métrica TF-IDF asigna a cada palabra un peso de relevancia en el corpus, el cual denotamos como $\omega_j(t_i)$, donde j hace referencia al j-ésimo evento focal y t_i se refiere al i-ésimo texto en el corpus. De tal forma que tenemos un peso para cada elemento en $\mathcal{L}_\Omega$ por cada corpus en el que esta presente, en este caso en particular 3 pesos para cada palabra en el léxico, uno por cada evento focal analizado. Con el objetivo de resumir la relevancia de cada palabra en $\mathcal{L}_\Omega$, definimos el peso total del texto $t(i)$, denotado por $W(t_i)$, como:

$$W(t_i) = \sum_{j=1}^{N} \omega_j(t_i). \tag{2.4}$$

Esta métrica es la más sencilla que puede definirse y es la que utilizamos para seleccionar a las palabras más relevantes dentro del léxico $\mathcal{L}_\Omega$.

Siguiendo este procedimiento construimos un léxico piloto compuesto de 365 palabras, de las cuales mostramos las 30 con mayor peso total en la Tabla 2.5 [8], es posible observar que en esta lista, además de palabras con connotación negativa de uso general, aparecen otras que corresponden al contexto social y político compartido por los distintos eventos focales analizados y que no necesariamente tienen una connotación negativa intrínseca, como por ejemplo: "morena", "calderón", "méxico", "madre", "abrazos", entre otros. Por lo que aún es necesario realizar un filtrado adicional para solamente considerar las palabras de uso general que no están relacionadas con un contexto particular.

Si bien el léxico obtenido y el procedimiento descrito anteriormente es perfectible, hemos obtenido un conjunto de palabras con connotación negativa no presentes en el corpus general TASS y que son representativas de las expresiones locales alrededor de los eventos focales analizados.

[8] *El léxico completo obtenido de este análisis se puede consultar en el repositorio de datos del Tlatelolco-Lab*

Palabra	Peso Total
duda	134.28
morena	124.87
pinche	116.82
calderón	92.14
méxico	86.26
madre	79.73
vale	66.45
...	60.51
abrazos	54.57
pendejo	52.28
jajaja	52.26
alguien	50.67
realidad	47.82
mexico	47.74
fracaso	47.05
asesinos	45.64
balazos	45.23
amlo	44.05
narco	43.3
dura	42.33
historia	42.06
lamentable	40.48
siguen	35.89
renuncia	35.05
pobre	34.63
neta	34.01
partido	33.23
asesino	33.1
mamá	33.01
gente	32.71

Tabla 2.5: Palabras con mayor peso total, $W(t_i)$ en el léxico $\mathcal{L}_\Omega$. Se observan palabras de uso general y otras que están relacionadas con el contexto político y social compartido por los distintos eventos focales a partir de los que se construyó el léxico.

3. Incorporando los hashtags al análisis de sentimiento

Los hashtags o etiquetas, son palabras o frases acompañadas del símbolo #, que pueden acompañar el texto de las publicaciones en Twitter y que son utilizados principalmente para agruparlos y relacionarlos con un tema o personaje con un sentido específico, en la conversación alrededor de un evento focal.

Si bien hemos observado que la inclusión de textos con expresiones y léxico local tienen un impacto en la clasificación que realizan los modelos de AS, también hemos identificado que los hashtags en los textos de las publicaciones (o tweets), aunque presentes en los corpus de entrenamiento, no tienen efecto en el proceso de clasificación realizado por los modelos de AS. Esto se debe principalmente a que, durante el proceso de entrenamiento del modelo de AS se genera un léxico con las palabras presentes en el corpus, incluidos los hashtags, en el que es poco probable que se encuentren los relacionados con temas actuales o coyunturales, ya que éstos se generan y mutan continuamente y por lo tanto el modelo de clasificación no cuenta con una referencia para evaluarlos. Sin embargo, las palabras que componen a los hashtags si pueden estar presentes en el corpus de entrenamiento y son susceptibles de ser aprovechados para complementar la clasificación del texto que hacen los modelos.

En específico, para intentar resolver esta problemática e incorporar a las palabras que componen a los hashtags en el proceso de clasificación, en primer lugar eliminamos el símbolo # y, en caso de que el hashtag se componga de varias palabras, usualmente separadas por letras mayúsculas, lo descomponemos, normalizamos e integramos en el texto original de la publicación que será evaluado por el modelo. Es decir, se siguen los siguientes pasos:

- Eliminamos el símbolo # y convertimos toda mayúscula a minúscula dando como resultado:
  #Narco es remplazado por **narco**
- Separamos los hashtags compuestos localizando las mayúsculas de la manera siguiente:
  #MorenaNarcoPartido es remplazado por **morena narco partido**

3.1. Resultados

En la Tabla 3.1 podemos observar que algunos de los tweets clasificados con alta probabilidad de pertenecer a la categoría positiva, p^+, por el modelo entrenado con el corpus con ejemplos de español de España, no considera los hashtags para su evaluación, por ejemplo textos con los hashtags: #MorenaNarcoPartido o #NarcoPresidente, que incluyen la palabra “narco” que puede ser usada con una connotación negativa.

Texto	Positivo	Negativo
”@TraducSerPRO Muchas gracias Pedro por informarnos, muchas felicidades por tu trabajo y aporte a la niñez mexicana.#NiUnVotoAMorena2022 #NarcoPresidente @AccionNacional @PRDMexico @PRI_Nacional	0.999820	0.0001665
Y el CO feliz, feliz, feliz!! #NarcoPresidente https://t.co/56nJ4y8dN	0.999606	0.00044265
Abrazos por todos lados... #NarcoEstado #NarcoPresidente	0.999467	0.000553733
Gracias al #MorenaNarcoPartido #NarcoPacto #NarcoPresidente	0.999393	0.000667777

Tabla 3.1: Ejemplos de textos de Tweets que utilizaron #NarcoPresidentecon con los valores más altos de p^+ asignados por el modelo de AS entrenado con el corpus general TASS

Como se muestra en la Fig. 3.1, la incorporación de los hashtags al texto por medio de la estrategia descrita anteriormente, puede generar un cambio en su clasificación en ambas direcciones, es decir, pueden pasar de positivos a negativos o viceversa. Sin embargo, puede también darse el caso que, como lo describimos en la sección 2.4 en el caso de los corpus extendidos con expresiones locales, las palabras que componen los hashtags no se encuentren dentro del léxico del corpus de entrenamiento del modelo, y por lo tanto no contribuyan a su clasificación (como los puntos alrededor de la identidad en la Fig. 3.1), o que el cambio de clasificación no corresponda con el resultado esperado debido a que no existen ejemplos en el corpus de entrenamiento con el mismo contexto, por ejemplo el caso de la palabra “narco” mencionado anteriormente.

Así mismo, como mostramos en la Tabla 3.2, y dado que la incorporación de los hashtags en el texto a ser evaluado puede generar un cambio de clasificación en cualquier dirección, podemos observar que puede darse el caso en el que el contraste en el sentimiento del texto y el de los hashtags pueda ser tal, que al incorporarlos en la evaluación del texto se obtenga la clasificación contraria del texto.

Una manera de analizar e identificar este contraste, es evaluando a cada componente por separado, es decir presentar por separado el texto y el resultado de expandir los hashtags que lo acompañan al modelo y obtener sus probabilidades p^- y p^+ correspondientes y compararlos. En la Tabla 3.3 mostramos algunos ejemplos en los que la clasificación del texto y de los hashtagas es contraria.

Como se puede observar, hay casos en los que algunas personas usuarias de la plataforma acompañan los textos de sus publicaciones con hashtags como: #FelizLunes, #FelizViernes, que tienen un sentido positivo, pero que al mismo tiempo se expresan de manera negativa con respecto a algún actor político o social. Este primer caso puede no ser representativo ya que puede darse el caso de que el usuario utilice un hashtag popular o

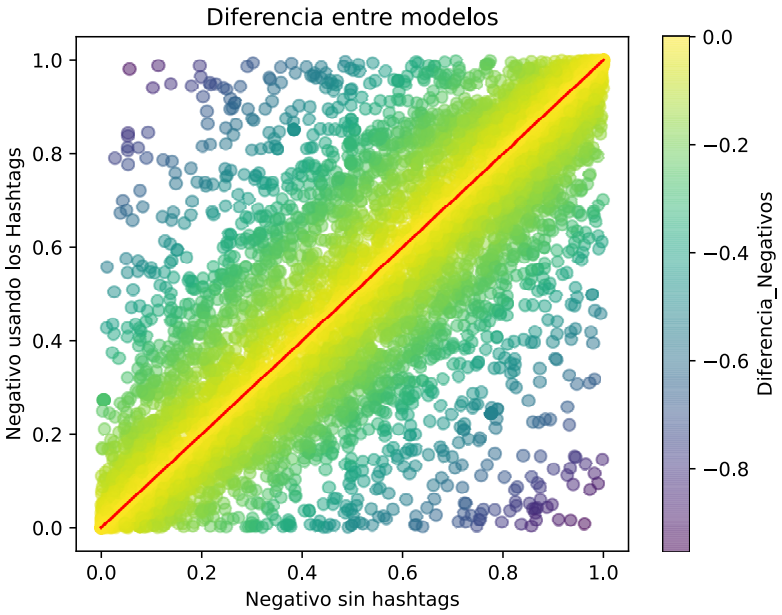

Figura 3.1: Cambio en la evaluación de sentimiento de los tweets de las elecciones presidenciales en México para los textos sin alterar y los textos con los hashtags modificados, cada punto representa un tweet y el color representa la diferencia entre modelos entre mas oscuro mayor la diferencia.

con gran alcance para difundir su verdadero mensaje, práctica usualmente conocida como "secuestro de hashtag". Otra posibilidad es aquella en la que el hashtag si tiene una intención, como en el caso del tercer ejemplo de la Tabla 3.3, donde el usuario se comunica con otro de manera positiva pero los hashtags que utiliza tienen claramente el sentido contrario.

Este ejercicio, aunque no concluyente por el momento, puede ser una vía para la identificación y análisis del uso sarcástico o irónico de hashtags en la conversación que se desarrolla alrededor de eventos focales en las redes socio-digitales en un contexto social y político.

Texto	Negativo	Positivo
Texto original: #BuenasTardes #FelizLunes Así de Metida está la delincuencia organizada en el Gobierno de #AmloPerroDelNarco ..🤬 https://t.co/RKtIDhphMi	0.961372	0.0438935
Texto modificado: buenas tardes feliz lunes Así de Metida está la delincuencia organizada en el Gobierno de amlo perro del narco ..🤬 https://t.co/RKtIDhphMi	0.0087301	0.990831
Texto original: Tienen esta cantimplora y la horrible... #FelizMartesATodos #EstadoDeExcepcion #AmberHeard #AmloPerroDelNarco https://t.co/zCYu2b1Qc6	0.934406	0.0654737
Texto modificado: Tienen esta cantimplora y la horrible... feliz martes atodos estado de excepcion amber heard amlo perro del narco https://t.co/zCYu2b1Qc6	0.0161235	0.979709
Texto original: #Hidalgo piensa tu VOTO ¿Cómo está tu cartera, seguridad, servicios (agua,luz,gas) Cuidado con #ElPeorGobiernoDeLaHistoria https://t.co/Ewsf5BJLej	0.102595	0.907872
Texto modificado: hidalgo piensa tu VOTO ¿Cómo está tu cartera, seguridad, servicios (agua,luz,gas) Cuidado con el peor gobierno de la historia https://t.co/Ewsf5BJLej	0.94068	0.0848953
Texto original: Ojalá que la #OposicionTotalmenteDerrotada #OposicionMiserableMezquinaYTraidora haya aprendido que los #BotsNoVotan 🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬	0.138171	0.861773
Texto modificado: Ojalá que la oposicion totalmente derrotada oposicion miserable mezquina y traidora haya aprendido que los bots no votan 🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬🤬	0.948845	0.054351

Tabla 3.2: Ejemplos de comparación de las probabilidades de pertenencia asignadas por el modelo de AS al incorporar los hashtags en el texto a ser evaluado.

Texto	Negativo texto	Positivo texto	Negativo hashtag	Positivo hashtag
La Cuarta transformación de #Veracruz y su voracidad por el Dinero acabará con el #AcuarioDeVeracruz #CuitlahuacGarcia a la cárcel #FelizMiercoles #AMLOamigodeNarcos https://t.co/EAnuabxXBi	0.99841	0.00110	0.00049	0.99965
Una vez más, @lopezobrador_ fue traicionado por su subconsciente, y dijo "Un gobierno sin corrupción no sirve para nada... Y eso es la #4T #FelizMiercoles @Wendympinto @CNNEE Loco de Atar... https://t.co/rJSKleElfh	0.99753	0.00288	0.00043	0.99969
@MarkDeReborn Aspiracionista... (No es cierto, enhorabuena) Salud caballero, increíble la resistencia... Somos más, los que no queremos a #ElPeorGobiernoDeLaHistoria #NarcoGobierno Cuánto gana #JoseRamonLopezBeltran111 ?	0.00211	0.99761	0.99396	0.00575
Muchas gracias #NiPerdónNiOlvido #6De6ParaMorena #5DeJunioNoSeOlvida 49 angelitos #GuarderíaABC13años https://t.co/oj41SvNYks https://t.co/1wLIMRIAv1	0.00104	0.99920	0.99104	0.00810

Tabla 3.3: Ejemplos de comparación de la evaluación del texto y de los hashtags expandidos por separado por el modelo entrenado con el corpus general TASS. Se puede observar como hay casos en donde se tienen evaluaciones contrarias.

4. Conclusiones

Como hemos mostrado a lo largo de este documento, las características del corpus de entrenamiento de los modelos de AS, específicamente de aquellos basados en RNA como el que utilizamos en este análisis, tienen un papel relevante en la clasificación de los textos que se le presenten. En particular observamos que el corpus general TASS, compuesto por textos escritos en español de España, no es suficiente para identificar casos de textos negativos que incluyen palabras o expresiones locales características.

Tanto en el caso de México como en el de Colombia, observamos un cambio de clasificación, es decir los textos clasificados pasan de ser positivos a negativos y viceversa, cuando son evaluados con modelos de AS entrenados con corpus extendidos que incluyen expresiones locales y del contexto particular de cada región. Esta diferencia entre la clasificación obtenida del modelo entrenado con el corpus general y los modelos con los corpus extendidos, ponen de manifiesto la necesidad de contar con un corpus anotado (es decir, cada elemento del corpus está clasificado como positivo o negativo previamente) actual y que refleje de mejor manera el lenguaje de la región. Si bien ésta es una tarea complicada, es susceptible de ser implementada mediante el uso de plataformas en línea y en colaboración con la ciudadanía. Esta será una vía que exploraremos en un futuro.

Así mismo, hemos identificado que los hashtags también son relevantes para la clasificación de textos realizada por modelos de AS, ya que, de no estar presentes en el corpus de entrenamiento del modelo, estos no contribuyen en la clasificación que se realiza de los textos. Como ya hemos mencionado previamente, los hashtags utilizados en las redes socio-digitales, y en particular en Twitter, pueden tener una vida corta, generarse y mutar rápidamente, por lo que hace complicado que los hashtags actuales estén presentes en los corpus de entrenamiento que han sido desarrollado hace algunos años. La estrategia de incorporación de hashtags que proponemos en este documento, puede ser una vía de para la inclusión de los hashtags en la evaluación de los tweets, sin embargo, se presenta el mismo problema de contar con un corpus que represente los términos y el lenguaje utilizados cotidianamente en la plataforma.

Por otro lado, también hemos mostrado que otra manera de incorporar la información contenida en los hashtags, es mediante la evaluación independiente del texto y de las palabras que los componen. El contraste en la evaluación de ambos elementos puede ser una estrategia para identificar el uso sarcástico o irónico de los hashtags al identificar casos en los que las evaluaciones del texto y los hashtags son opuestas.

Los resultados obtenidos en este documento, muestran distintas líneas de investigación para explorar con mayor profundidad y desde distintos ángulos, el análisis de sentimiento de los textos que se comparten en las redes socio-digitales.

Bibliografía

[1] Julio Villena Román et al. ***TASS - Workshop on Sentiment Analysis at SEPLN***. 2013.

[2] S. Robertson. «Understanding inverse document frequency: on theoretical arguments for IDF». En: ***Journal of Documentation*** Vol. 60 (2004), págs. 503-520.

[3] Martín Zumaya, Diego Espitia y Luis Ángel Escobar. ***Análisis de Sentimiento en Twitter. Aproximaciones con métodos de Aprendizaje de Máquinas***. Documento de trabajo. PUEDJS-UNAM, México, 2022.

[4] Martín Zumaya, Diego Espitia y Luis Ángel Escobar. ***Comunidades en redes socio-digitales. Identificación de narrativas y cuentas representativas***. Documento de trabajo. PUEDJS-UNAM, México, 2022.

El presente documento es producto de una investigación realizada en el marco de los Programas Nacionales Estratégicos del Consejo Nacional de Ciencia y Tecnología (PRONACES-CONACYT). Agradecemos al CONACYT por el generoso apoyo brindado en 2022. El contenido y las opiniones son responsabilidad exclusiva de los autores.
| **contacto: contacto.puedjs@gmail.com**