

Un análisis de sentimiento en Twitter con Machine Learning: Identificando percepciones sobre el Día Sin IVA.

La técnica de análisis de sentimientos se define como el campo de estudio que analiza las opiniones de las personas, sentimientos, apreciaciones, actitudes y emociones sobre algún evento en particular; así, esta herramienta se ha constituido en una recurrente aplicación de la lingüística computacional. Adicionalmente, la red social Twitter se ha convertido en un importante instrumento para el análisis de audiencias y el estudio de opiniones sobre eventos y acontecimientos en tiempo real alrededor del mundo.

En este contexto, el área de **Estudios Económicos y Competitividad de la Cámara de Comercio de Manizales por Caldas** empleando una metodología de Análisis de Sentimiento con Machine Learning busca identificar el sentimiento y percepción de los usuarios sobre la medida del día sin IVA en el periodo de tiempo del 18 de mayo al 03 de julio de 2020.

Contexto del día sin IVA

El día 27 de diciembre de 2019 se sancionó la Ley 2010 de 2019 o Ley de Crecimiento Económico. La Ley 2010 está enfocada a estimular el desarrollo económico y sentar bases para reducir la desigualdad y cerrar sus brechas. Adicional, se definen temas de tipo social, como la devolución del IVA para 2,8 millones de hogares de escasos recursos, mecanismo efectuado en el marco de la emergencia sanitaria causada por el Covid-19. En continuidad, se trata la reducción de cotización en salud del 12 al 4% para los pensionados más vulnerables e incentivos a las empresas que creen puestos de trabajo para jóvenes de entre 18 y 28 años. En este contexto se definió, además de aumentar recursos destinados a educación, salud, mejoramiento de viviendas y llevar acueducto a las zonas más vulnerables incluir dentro del calendario tres días sin IVA al año.

Con el objetivo de reactivar la economía local debido a la crisis generada por el covid-19, las fechas sin IVA definidos por la Dirección de Impuestos y Aduanas Nacionales DIAN fueron el 19 de junio, 3 de julio y 19 de julio. Según la reglamentación aprobada en la Ley 2010 de 2019 y los ajustes realizados en el marco de la emergencia sanitaria, la medida cubre las siguientes categorías: vestuario, accesorios, útiles, electrodomésticos, equipos deportivos, las patinetas y bicicletas, tanto mecánicas como eléctricas, computadores,

celulares y equipos de comunicación e insumos agropecuarios. Finalmente, los productos con un valor hasta \$2,8 millones estarán exentos del cobro de este impuesto.

Metodología

Este ejercicio emplea la red social Twitter como objeto de estudio. Lo anterior dado que esta se ha constituido en un canal de uso común por millones de usuarios los cuales buscan interactuar de forma simultánea y continua (Kuo et al., 2012). En continuidad, la cantidad de datos vinculados a un evento en particular desde Twitter pueden ser segmentados por la ubicación geográfica y recopiladas en función de una etiqueta (numeral #hashtag). Por otra parte, y en su generalidad, el análisis de sentimiento tiene implícito dos objetivos. Primero, reconocer las expresiones de sentimiento y segundo, definir la orientación del sentimiento expresado por los individuos.

En este caso, el algoritmo computacional permite, a partir de la etiqueta #DíaSinIva, recopilar los mensajes que se publican en la red social desde Manizales. De esta forma, a partir del Hashtag sobre los tweets de los diferentes usuarios se recopiló información respecto al contenido señalado.

Tabla 1. Descripción de las emociones

Dimensiones	Variables
Enfado	Ira
	Fastidio
Anticipación	Anticipación
Miedo	Temor
	Horror
	Pánico
Alegría	Júbilo
	Regocijo
Tristeza	Resignación
	Desilusión
Sorpresa	Asombro
	Perplejidad
	Confusión
Confianza	Equilibrio
	Serenidad
	Tranquilidad
Disgusto	Disgusto

Fuente: Estudios Económicos y Competitividad CCMPC.

La estrategia definida para recopilar la información está fundamentada en las siguientes etapas. Primero, la conexión a la API de Twitter desde la cual se ha descargado un total de $n=5000$ tweets procedentes de los comentarios e interacciones que realizan los usuarios sobre la etiqueta #DíaSinIvra en el periodo del 18 de mayo de 2020 hasta el 03 de julio del mismo año. En segunda instancia, para la aplicación de algoritmo desarrollado en R, se ha empleado la librería *syuzhet*; a partir de allí se ha aplicado el algoritmo de análisis de sentimiento para dividir los tweets en positivos, negativos y en 8 categorías de emociones, tal y como se muestran en la tabla 1. Finalmente, se incorporan herramientas del análisis de texto como las nubes de palabras para extraer información descriptiva sobre las emociones.

Resultados

De acuerdo con la información que se presenta en la Figura 1, las palabras “comprar”, “cumplimiento” y “bioseguridad” reflejan una mayor frecuencia en los mensajes publicados en la red social Twitter. Lo anterior si bien no sorprende, permite inferir cierto grado de responsabilidad en relación con la emergencia sanitaria y el propósito de reactivar la economía.

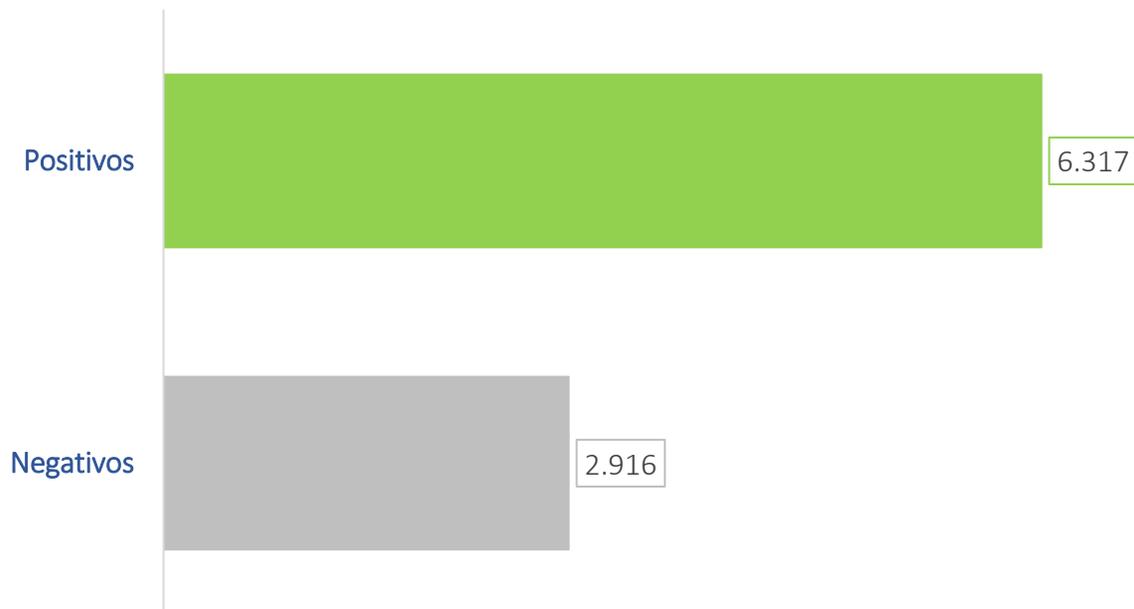
Figura 1. Nube de palabras



Fuente: Estudios Económicos y Competitividad CCMPC.

Por otra parte, tras el desarrollo del proceso metodológico que incluye la recolección de los datos y la extracción de estos, se ha procedido a realizar el análisis de sentimientos. En la Figura 2 se pueden observar los resultados del proceso en el que se identifican el número de palabras procesadas en los Tweets con respecto al uso del hashtag #DíaSinIva por parte de los usuarios en la ciudad de Manizales y las interacciones y comentarios que realizan en ellos.

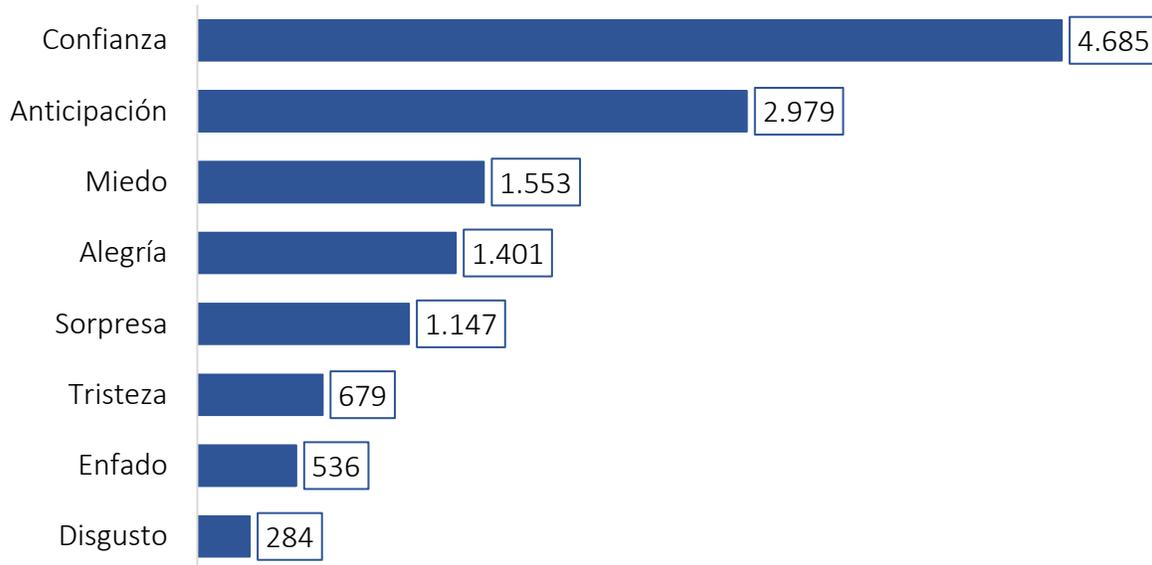
Figura 2. Número de palabras categorizadas según sentimientos.



Fuente: Estudios Económicos y Competitividad CCMPC.

Dado que el número de palabras categorizadas como negativas constituye una tercera parte del total de términos contrapuestos, los resultados del análisis permiten inferir consideraciones favorables de los usuarios de Twitter frente a la medida del Día Sin Iva. En continuidad, estos hallazgos vislumbran los procesos de preparación conjunta efectuados por alcaldes y sector empresarial con el fin de disminuir las aglomeraciones presentadas el día 19 de junio.

Figura 3. Número de palabras categorizadas según emociones.



Fuente: Estudios Económicos y Competitividad CCMPC.

En análisis de los resultados, es plausible que un porcentaje de las palabras empleadas en los tweets categorizados como Enfado o Disgusto obedezcan a las dificultades presentadas por plataformas online a raíz la congestión de la red informática. En esta línea, la tercera emoción en relación con la frecuencia es “Miedo”; como es de esperar, los fenómenos de aglomeración derivados por la medida generan temor y pánico, según reflejan los comentarios y opiniones de los usuarios en Twitter.

En contraste, la emoción con mayor de frecuencia dentro los 5000 tweets analizados es “Confianza”. Este hecho refleja, por un lado, cambio en la percepción respecto a la autodisciplina y las lecciones aprendidas del primer día sin IVA en el país. Por otra parte, este hallazgo revela, en general, mejor adecuación de las medidas de bioseguridad promovidas por alcaldes y efectuado esencialmente por las grandes superficies.

Conclusiones

De la forma en la que se ha expuesto, Twitter es una red social óptima para que los usuarios y audiencias puedan expresar sus sentimientos, opiniones y comentarios sobre un tema concreto, en tiempo real y alrededor del mundo. En particular, el área de **Estudios Económicos y Competitividad de la Cámara de Comercio de Manizales por Caldas** ha empleado Twitter para realizar una clasificación del sentimiento de los usuarios utilizando el hashtag #DíaSinIva desde el 18 de mayo de 2020, día en el cual se anunció los 3 días en los cuales los bienes estarían exentos de IVA hasta el 03 de julio de 2020.

Los resultados de la investigación descubren aquellas comunicaciones realizadas en Twitter sobre las ideas que los usuarios han expresado como positivo o negativo. En este sentido, podemos concluir que el 68% de los tweets categorizados son clasificados como positivos, de lo cual se concluye que en la ciudad de Manizales los usuarios respaldan la medida adoptada en el marco de las políticas de reactivación económica. Entre tanto, la nube de palabras, la cual se presenta con fines descriptivos, permite inferir mejor preparación y coordinación por parte de los alcaldes locales y sector privado en el segundo día sin IVA.

Equipo de investigación

Mateo Andrés Rivera Arbeláez
Estudios Económicos y Competitividad
mrivera@ccm.org.co

Jorge Alberto García García
Estudios Económicos y Competitividad
jgarcia@ccm.org.co

Juan Manuel Aristizábal Tamayo
Investigador asociado
jmaristizabal@uniquindio.edu.co