

Eficacia de la red social Twitter en la detección de casos de congestión vehicular

Efficiency of the social network Twitter in the detection of cases of vehicular congestion

Nelson Herrera Herrera¹

nelson.herrera@ute.edu.ec

<https://orcid.org/0000-0002-5781-6444>

Estevan Gómez Torres²

ergomez@espe.edu.ec

<https://orcid.org/0000-0002-1171-7256>

Recibido: 2/8/2022; Aceptado: 20/6/2023

RESUMEN

El trabajo tiene como finalidad demostrar la eficacia del uso de la red social Twitter en la detección en tiempo real de situaciones de congestión vehicular, considerando como caso de estudio a la ciudad de Quito-Ecuador. Para lo cual se utilizó el API (Application Programming Interface) de la red social que recoge datos de posibles puntos de congestión vehicular, mediante herramientas de Big Data. Los datos obtenidos son correlacionados con información obtenida mediante sensores de flujo vehicular, para su posterior análisis lo cual permitió identificar situaciones de congestión vehicular.

Palabras clave: congestión, sensores, Twitter, API, Big Data.

ABSTRACT

The purpose of the work is to demonstrate the effectiveness of the use of the social network Twitter in the real-time detection of traffic congestion situations, considering the city of Quito-Ecuador as a case study. For which the API (Application Programming Interface) of the social network was used that collects data from possible points of vehicular congestion, through Big Data tools. The data obtained is correlated with information obtained through vehicular flow sensors, for subsequent analysis, which allowed the identification of situations of vehicular congestion.

Keywords: congestion, sensors, Twitter, API, Big Data.

¹ Doctor en Modelización y Experimentación de Ciencia y Tecnología, Universidad UTE, Ecuador

² Master en Gerencia de Sistemas, Universidad de las Fuerzas Armadas, ESPE, Ecuador

Introducción

Hoy en día un alto porcentaje de ciudades del mundo, están utilizando soluciones tecnológicas que contribuyan a mejorar la calidad de vida de sus ciudadanos, incursionando en el concepto de Ciudades Inteligentes, y la gestión de tráfico inteligente en la ciudades (Digi, 2022), que corresponde a ciudades en las que la implementación de tecnologías de la información y comunicación, es considerada como una excelente alternativa y tomando como referencia su análisis y aplicación, buscan la eficiencia del desarrollo urbano, basado en la sostenibilidad que deberá ser capaz de responder de manera eficiente a las necesidades actuales.

Uno de los problemas principales en algunas ciudades a nivel mundial es la "Congestión Vehicular" y la ciudad de Quito no es la excepción. Debido a la expansión y desarrollo de la ciudad, motivado por el boom de la actividad comercial, ha aumentado la movilidad de los ciudadanos en la ciudad. Para cubrir esa necesidad, la cantidad de vehículos ha ido incrementándose en los últimos años, que unido a la expansión relativamente lenta de las carreteras hacen que la congestión vehicular se convierta en un problema grave (Quito M. d., 2014) .

Cada vez son más frecuentes los embotellamientos de tráfico y los ciudadanos emplean más tiempo en trasladarse de un lugar a otro, principalmente en horas pico (Quito A. , 2011). Además, la topografía de la ciudad, con grandes montañas al lado occidental y desniveles y valles al lado oriental ha generado que la misma se expanda en dirección norte – sur, es decir de forma alargada y poco organizada y planificada, agravando el problema.

La congestión del tráfico no solo genera pérdida de tiempo y dinero, también aumenta la contaminación del aire debido a la emisión de gases adicionales. Otros efectos de la congestión del tráfico incluyen: obstaculizar los servicios de emergencia - como la policía y los servicios médicos y aumentar el vehículo se bloquea cuando los automovilistas se apresuran para evitar quedar atrapados en un embotellamiento (Herrera, 2016).

En la actualidad, en la ciudad de Quito, se han aplicado varias alternativas y medidas que pretenden mejorar la movilidad en la ciudad, un ejemplo es el uso de sensores, semáforos inteligentes, entre otras (Pozo, 2016). Se han planteado soluciones IoT para la detección de espacios libres en parqueaderos, (Avalos & Gomez, 2019). Sin embargo, los resultados de aplicar estas medidas no han dado los resultados esperados, los problemas de tráfico persisten de manera permanente, además de que no todos los planes se han aplicado en su totalidad.

Se han realizado estudios en todo el mundo, con respecto a soluciones de congestión vehicular (IESE, 2017) , uno de ellos es el realizado por los investigadores del Instituto Tecnológico de Massachussets (MIT) y la Universidad de Birmingham, en el cual se realizó un plan piloto en cinco ciudades, en donde mediante el uso de teléfonos inteligentes reorientó a los conductores, buscando optimizar rutas y flujos alternativos. La investigación mostró que con este tipo de aplicación se podría reducir un 30% el tráfico (Delgado, 2016). Si bien este resultado es de una aplicación local, se puede

Herrera, Gómez.

Eficacia de la red social Twitter en la detección de casos de congestión vehicular

inferir este resultado a otros casos, ciudades, así como a otras aplicaciones, es así como las soluciones pueden ser planteadas y aplicadas a la ciudad de Quito.

El desarrollo de la informática y las telecomunicaciones trae consigo además un incremento de la cantidad de dispositivos, celulares y equipos electrónicos conectados a redes de comunicaciones, y es común que en cada vehículo pueda haber pasajeros conectados a redes sociales como Facebook, Instagram y Twitter. Producto del largo tiempo que pasan estos pasajeros en los embotellamientos puede traer consigo un aumento de la interacción con estas redes sociales y muchas veces se pueden referir en sus textos a los mismos embotellamientos.

El presente trabajo tiene como objetivo presentar una alternativa eficiente a los problemas de congestión vehicular en la ciudad de Quito con un menor costo y mayor cobertura espacial utilizando la red social Twitter como fuente de datos. Utilizando la correlación entre los datos obtenidos por sensores y los datos capturados por los Tuits se describen modelos que nos permiten estimar situaciones de congestión vehicular.

Metodología

Para el cumplimiento de los objetivos, la metodología de trabajo se presenta a través de las fases, que se muestran en la Figura 1:

Fase de generación

Figura 1.
Utilizadas
Fuente:



Fases

Autores

La fase de generación contempla la entrada de los datos. Las fuentes generadoras son: Red Social Twitter: para la recolección de tuits se creó el hashtag #TraficoUIO y se utilizó las siguientes palabras clave (Keywords): atasco, embotellamiento, tránsito, circulación, congestión, atascamiento, tráfico; además de utilizar los siguientes hashtags creados por el CGM: #MovilidadUIO, #ViasAlternasUIO, #RadaresUIO.

Sensores de volumen de tráfico: ubicados en ciertas intersecciones consideradas conflictivas para la congestión vehicular. La información generada por estos dispositivos, corresponde a la cantidad de autos que están circulando a una distancia promedio de tres metros, a partir de la ubicación del sensor instalado hacia adelante. Sensores con circuito cerrado de televisión (técnicas de imagen): son desplegados junto a los semáforos ubicados en toda la ciudad, incluidos sectores en los cuales no están desplegados los sensores de volumen de tráfico.

Fase de recolección

Sensores de flujo vehicular

La primera componente está relacionada con la información recolectada por los sensores que alimentan el Sistema de Registro Vehicular, que es manejado por el Centro de Gestión de la Movilidad (CGM) de la ciudad de Quito, en el mismo se registran la carga vehicular existente en determinadas intersecciones de la ciudad consideradas como conflictivas para la congestión vehicular. La información generada y almacenada en un FTP en formato xlsx; describen información relacionada intersecciones, puntos georreferenciales (coordenadas), fecha, hora y umbrales de congestión.

Sensores con circuito cerrado de televisión

Estos dispositivos están conectados en todas las intersecciones de la ciudad, incluyendo aquellas intersecciones en las cuales no están instalados, sensores de flujo vehicular. Las cámaras de CCTV situadas estratégicamente pueden ser visionadas desde el CGM de la ciudad en un video-wall de 36 módulos de 50"; el nivel de congestión se mide mediante el procesamiento de imágenes. Mediante un software encargado de procesar cada una de las imágenes obtenidas y posteriormente realizar el reconocimiento y conteo de vehículos.

Red Social Twitter

Inicialmente se recolectó todos los tuits asociados al hashtag antes mencionado haciendo uso del API (Application program interface) de Twitter. La configuración del tuit considera la ubicación desde la que se genera el mismo con lo cual es posible obtener la localización del tráfico, para los tuits, que denuncien esta situación.

La información procedente de las redes sociales fue recopilada con la herramienta Apache Flume. Esta herramienta es de gran utilidad debido a que permite recolectar y almacenar los tuits generados por los hashtags y las keywords, en el sistema de archivos Hadoop Distributed File System de Apache Hadoop. La Figura II muestra los componentes del agente Apache Flume.

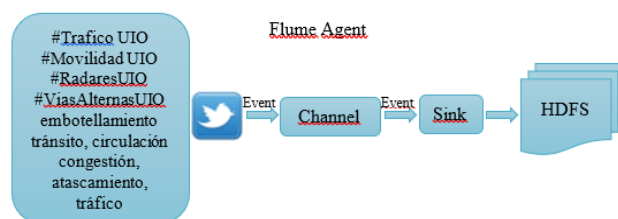


Figure 2. Colección de Datos Apache Flume

Fuente: Elaboración propia

Fase de integración

La información en bruto procedente de la red social Twitter, se la debe filtrar con la finalidad de eliminar información no relacionada directamente con temas de congestión vehicular, para lo cual se aplican técnicas de Análisis de Sentimientos que se detallan en la siguiente sección. Posterior a estos análisis la información resultante es trasladada al repositorio Apache HBase, mediante la herramienta Apache Mahout. Por otra parte, la información procedente de los archivos entregados por el CGM y depositados en el FTP (ftp:\\UIO\\CGM\\Transito), se conectan al repositorio Apache HBase, mediante un componente integrado (Excel Add-In), que se lo debe configurar en los archivos de Excel. Finalmente, para cargar la información de los archivos HDFS y los archivos de Excel, se creó la tabla Tweets en Apache Hbase, con los campos: 'intersección', 'coordenadas', 'hora', 'día', que almacenara la información obtenida de las diferentes fuentes mencionadas anteriormente. En la Figura 3 se puede apreciar el esquema utilizado para esta fase.

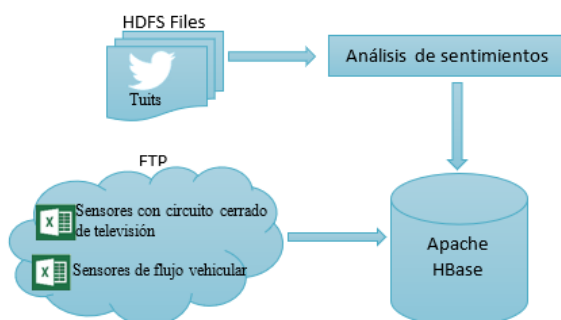


Figure 3. Information Integration Architecture

Fuente: Elaboración propia

Fase de análisis: La información de tuits recolectada y almacenada temporalmente en los archivos HDFS, debe ser filtrada antes de trasladarla finalmente al repositorio Apache HBase. Para lo cual se utiliza la herramienta del ecosistema Hadoop: Apache Mahout. En esencia es una biblioteca de Java y por ende no provee una interfaz de usuario o un instalador sino más bien es un conjunto de herramientas. Los algoritmos con los que trabaja Mahout se enfocan principalmente en tres áreas: sistemas de recomendación, agrupación y clasificación, (Owen, 2012).

La presente investigación centra sus estudios en los sistemas de clasificación (Wang, 2015), que son una forma de aprendizaje automatizado que utiliza algoritmos para proveer a las computadoras una forma de tomar decisiones en base a la experiencia, y de cierta forma emular el proceso de toma de decisiones del ser humano.

Para el análisis este estudio utiliza técnicas de Minería de Datos (Data Mining) de aprendizaje supervisado (Feldman, 2007), sobre los archivos HDFS que contienen información de los tuits recolectados. Este tipo de algoritmos consta de tres fases principales, la primera de recolección y construcción del Corpus, la segunda la extracción y selección de características y la tercera aplicación del algoritmo clasificador. (Zafarani, 2014)

Generación del corpus: Un Corpus es un conjunto de textos que son usualmente guardados y procesados por computadores y usados con fines estadísticos para contrastaciones de hipótesis y así validar reglas lingüísticas. En Análisis de Sentimientos, forma parte esencial de las aplicaciones y su desempeño, debido a que los textos contenidos en el Corpus conforman los datos de entrenamiento que se utilizarán para el aprendizaje de los algoritmos de clasificación. (Agarwal, 2011) Para los algoritmos de aprendizaje supervisado y categorización de clases, se deben construir Corpus que representen fehacientemente cada una de las clases para que luego de extraídas las propiedades de los textos contenidos en cada una de ellas, el algoritmo sea capaz de clasificar correctamente nuevos textos no procesados anteriormente.

Los tuits recolectados componen el Corpus, el cual está compuesto por los 109502 mensajes de tuits recolectados con la herramienta Apache Flume, se establecieron tres clases; Sentimientos Positivos, Negativos y Neutros. Para esto se utilizará el procedimiento que consiste en la recolección de tweets con emoticones (emoticons)

polares (Varghese Babu & Fabeela, 2021) teniendo en cuenta que un emoticon polar tiene una alta correlación con la polaridad del texto que lo contiene.

Extracción de características: El proceso de extracción de las características de los datos es uno de los más complejos ya que en base a estos se entrenan los algoritmos que posteriormente realizarán la clasificación. Existen 4 categorías de características que han sido utilizadas en trabajos previos. Estos incluyen características del tipo sintáctico, semántico, basado en links y de estilo. Entre estas características las más utilizadas son las semánticas y sintácticas para la tarea de Análisis de Sentimientos. Estas incluyen el uso de n-grams, POS-Tags, y signos de puntuación (Sankalp et al., 2017). En particular se utilizará Treetagger (Schmind, 2018) para las tareas de anotación de los datos y obtener los Pos-tags presentes en los tweets. TreeTagger es una herramienta para anotar texto, que permite realizar un análisis morfosintáctico, desarrollada en Pyton y que se integra fácilmente con Apache Hadoop.

Clasificación de polaridad de sentimientos: Las técnicas utilizadas para la detección de polaridad de sentimientos pueden ser clasificadas entre tres categorías (Hasan, 2018). Estas incluyen algoritmos de Machine Learning, análisis de links y enfoques relacionados con la asignación de puntajes. Muchos estudios se han enfocado en el uso de algoritmos de Machine Learning entre los que destacan Support Vector Machines, Clasificadores Bayesianos y Máxima entropía, siendo el Clasificador de Naive Bayes el más utilizado entre estos (Arcila-Calderón et al., 2017). En este trabajo se utilizó el clasificador de Naive Bayes, que está basado en el teorema de Bayes, descrita por la ecuación (1), que para el caso de este trabajo asigna a un tweet representado por un vector de características, la clase con la mayor probabilidad de pertenencia. (Matuschka, 2021).

$$P\left(\frac{C}{F}\right) = \frac{P(C) * P\left(\frac{F}{C}\right)}{P(F)} \quad (1)$$

La probabilidad $P\left(\frac{C}{F}\right)$ determina la probabilidad de que la palabra esté, dada una cierta clase, la cual es extraída directamente del set de entrenamiento, tal como se explicó anteriormente, donde las clases ya están determinadas y basándose en la frecuencia de las palabras se puede obtener la probabilidad correspondiente. En este caso, se asume que la probabilidad que ocurra una palabra es independiente de otra, de manera que sea más simple realizar el cálculo, con lo cual $P\left(\frac{C}{F}\right)$ es simplemente un factor de normalización. Si bien, esta es una asunción fuerte, los resultados obtenidos son buenos, superando el 65% de certeza en la mayoría de los casos, al ocupar dos categorías de clasificación (a favor y en contra). Generalmente, antes de ocupar Naive Bayes se obtiene el conjunto de entrenamiento con el uso del diccionario léxico, asegurando un conjunto lo más preciso posible. Por último, se pueden ocupar los hashtags para el paso previo, sin embargo, no siempre se obtiene un conjunto lo suficientemente grande.

Después de haber realizado los procedimientos de análisis antes mencionados, en donde de un total de 107478 tuis recolectados inicialmente, esta cantidad se redujo

a 85983 tuis resultantes, que serán utilizados para la correlación con la información de sensores de tráfico y de cámaras de circuito cerrado.

Fase de presentación : Con los datos almacenados y el procesamiento de los mismos el sistema presenta información gráficamente, mediante la aplicación Carto DB, que es un servicio integrado con Apache Hadoop, debido a que toma la información directamente de Hbase/Hadoop y la representa espacialmente. La información representada muestra la base de datos tanto de los lugares donde existe un sensor en Quito, así como la base de datos de tuits que indican la existencia de tráfico en la ciudad. Esta herramienta web permite apreciar cartográficamente los resultados del análisis y representa zonas de congestión vehicular.

Resultados y discusión

El estudio se realizó en tres periodos, el primer estudio fue realizado en los meses comprendidos entre abril a diciembre del año 2020, el segundo periodo de recolección se obtuvo entre el primer semestre: abril a junio del año 2021, y finalmente el tercer periodo de recolección está comprendido entre enero y marzo del 2022; producto de esta recolección de datos se obtuvo un total de 85983 tuits con información sobre tráfico en la ciudad, discriminando tanto la existencia de congestión, como su ubicación, con la cual a continuación se presenta el análisis de esta información y sus principales resultados derivados de la misma. Esta información ha sido procesada para determinar tráfico en la ciudad, detectando casos de congestión por zonas, intersección, hora y día de la semana. La tabla siguiente, muestra un resumen de la información obtenida por las dos principales fuentes de datos utilizadas en el presente estudio: sensores y tuits. En la misma se escribe que el total de tuits sobre tráfico obtenido es de 85983.

Por otra parte, se muestra el número de situaciones de congestión vehicular detectada por los sensores y la cantidad de tuits que reportan una situación de congestión vehicular, así se puede mencionar que 12069 situaciones de congestión no fueron detectadas por ningún tuit, de la misma manera se puede apreciar que dos tuits, por ejemplo, reportaron cada una de las 8952 situaciones de congestión vehicular, obteniéndose un total de 17904 tuits asociados a los casos de congestión mencionados. Finalmente podemos mencionar que se determinaron un total de 68305 casos de congestión vehicular, de los cuales 56236 fueron cubiertos por los tuits en sitios donde están instalados sensores de volumen de tráfico y sensores con circuito cerrado de televisión. Esto se muestra en la Tabla 1.

Tabla 1. Distribución de tuits sobre tráfico vehicular

	Número de tuits	Situaciones de congestión	Total, de tuits
Intersección con sensores	0	12069	0
	1	38778	38778
	2	8952	17904
	3	4723	14169
	4	3783	15132
	Total	68305	85983

Fuente: Elaboración propia

Análisis por día de la semana

Parroquia

En la Figura 4, se puede apreciar la distribución de casos de congestión vehicular detectados por tuits y sensores segmentados por parroquias, distribuidos entre todos los días de la semana. Se puede evidenciar claramente que los días de mayores casos de congestión vehicular detectado por los tuits son los lunes y martes con 11144 y 9617 casos; que representan el 19.4% y 16.8% del total de casos registrados en todas las parroquias. Los casos detectados por los sensores, no están distantes de las estadísticas detectadas por los tuits, registrándose un total de 13200 y 11633 casos de congestión de igual manera los días lunes y martes que son los de mayor afluencia vehicular, que en porcentaje representan el 17.3% y 17% del total de casos registrados en las parroquias consideradas para nuestro estudio.

Los fines de semana, fueron los días de menor cantidad de situaciones de congestión, detectado por tuits y sensores, situación bastante entendible, pues son días de descanso. Sin embargo, el sábado presenta un valor atípico en el Centro Histórico (1046 y 1202 casos registrados por tuits y sensores respectivamente), lo cual se debe a la gran cantidad de eventos sociales y culturales desarrollados en el sector.

Finalmente podemos concluir, que la parroquia que registra mayor situación de congestión vehicular es Ñaquito, resultado que confirma la realidad de este sector, pues es el área financiera más importante de la ciudad de Quito, contrario sucede con la parroquia La Mena que es la de menor afluencia vehicular.

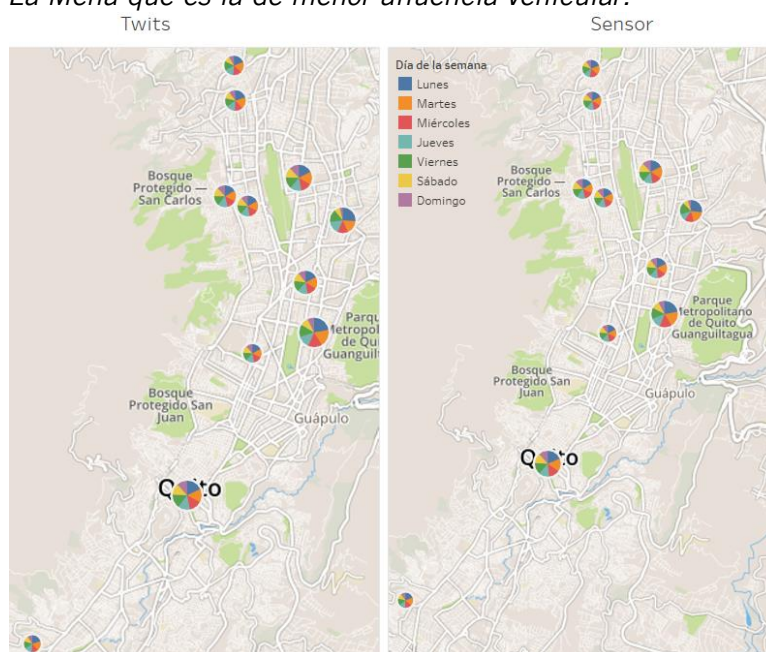


Figura 4. Situaciones de congestión detectadas por tuits y sensores en cada parroquia y día

Fuente: Elaboración propia

Correlación tuits y sensores

En la Figura 5, se puede apreciar la dispersión de tuits y sensores, que verifica una relación lineal dada por el modelo de tendencia calculado, siendo los valores de R^2 (coeficiente de correlación) valores muy cercanos a la unidad y el valor de $p < 0.05$

Herrera, Gómez.

Eficacia de la red social Twitter en la detección de casos de congestión vehicular

y por tanto valores significativos al 95% de confianza. La recta de tendencia generada para cada caso se ajusta de forma significativa a la recta $y=x$, donde y representa la cantidad de casos de congestión informados por los sensores y x representa los casos de congestión cubiertos por los tuits.

Se puede apreciar también que el día en el cual existe mayor porcentaje de cobertura es el jueves con un 92.7 % en una de las intersecciones de la misma parroquia, en donde existe el mayor porcentaje de cobertura, siendo el día sábado y domingo los días de menor cobertura con un 79.8%, en las intersecciones que pertenecen a la parroquia, que se ha identificado como la de menor porcentaje de cobertura.

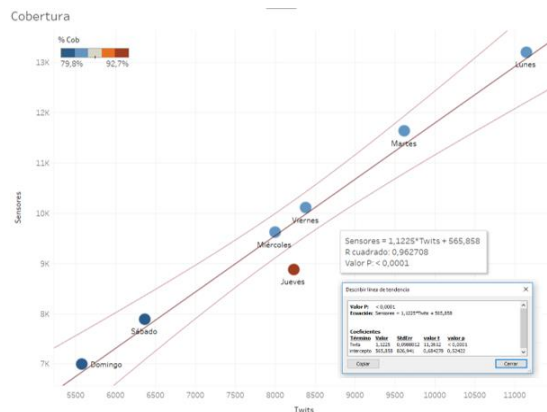


Figura 5. Correlación de tuits y sensores por día
Fuente: Elaboración propia

Intersección

La Figura 6, muestra la cantidad de casos de congestión vehicular por intersección y horas, pudiendo mencionar que existe mayor número de casos de congestión, detectada por la red social Twitter y sensores, en los horarios de 09:00 am a 10:00 am, 13:00 pm a 14:00 pm y en la tarde de 17:00 pm a 18:00 pm, además de mencionar que esta cantidad elevada de casos se detectaron en las intersecciones de las parroquias con mayor número de casos de congestión vehicular. Así por ejemplo se puede (Quito M. d., 2014) (Quito M. d., 2014) (Quito M. d., 2014) visualizar que el número de situaciones de congestión vehicular registrados por los sensores en el horario de 08:00 am a 09:00 am, es de 1015 casos en la intersección Vargas y Oriente perteneciente a la parroquia Centro Histórico, de un total de 4211 casos, que corresponde a un 24.1% del total de casos.

De igual forma se muestra la distribución de la cantidad de tuits registrados en el rango de horas comprendido de 07:00 am a 21:00 pm. Así podemos determinar, de la misma manera que se mostró en el caso del análisis por parroquia, que la mayor cantidad de tuits registrados están en los rangos de 07:00 am a 9:00 am, de 13:00 pm a 14:00 pm, y de 16:00 pm a 18:00 pm, los motivos son los indicados en el análisis de tuits de tráfico vehicular registrados, por intersecciones y horas, detallado en uno de los puntos anteriores y de igual forma las intersecciones en donde se registra la mayor cantidad de tuits de tráfico vehicular pertenecen a las parroquias indicadas como focos de tráfico vehicular, estas son: El Inca. Iñaquito y Centro Histórico.

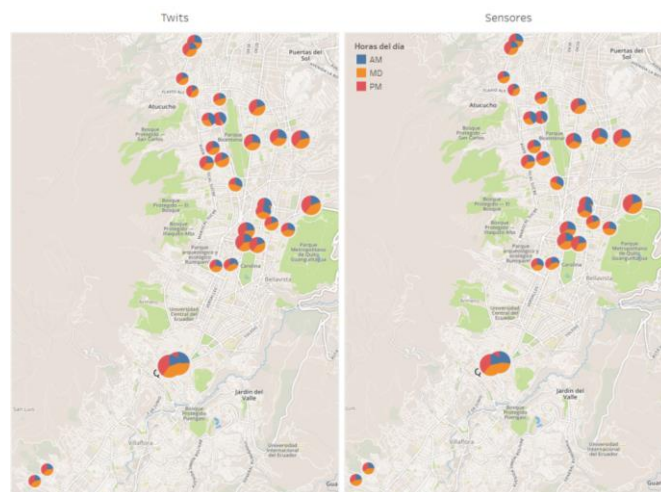


Figura 6. Situaciones de congestión detectadas por sen-sores y tuits por intersección y hora.
Fuente: Elaboración propia

Correlación tuits y sensores

Intersección

Con el objetivo de evaluar la correlación entre los casos de congestión vehicular registrados, por los sensores y los tuits generados, la información se presenta de igual forma mediante diagramas de dispersión para cada día de la semana. Para cada modelo obtenido podemos observar que se verifica una relación lineal dada por el modelo de tendencia calculado, siendo los valores de R^2 valores muy cercanos a la unidad y el valor de $p < 0.05$ y por tanto valores significativos al 95% de confianza.

La Figura 7, muestra el porcentaje de cobertura de los tuits en cada intersección en lugares en donde existen sensores de volumen de tráfico. Se puede mencionar que de igual manera que en el análisis de los días, el mayor porcentaje de cobertura se registró en la intersección García Moreno y José Joaquín de Olmedo en la parroquia Centro Histórico con un 89.6% y la de menor es cobertura Av. Antonio José de Sucre y Oe 10, perteneciente a la parroquia de Cotocollao con un 74.8 %, por cada hora comprendida en el intervalo de 07:00 am a 21:00 pm, la razón para tener estos resultados se los detallo en el análisis realizado por parroquia.

Haciendo el análisis por hora mencionamos también que la hora en la cual se registra un mayor porcentaje de cobertura es 17:00 pm a 18:00 pm y 19:00 pm a 20:00 pm, la primera se registra en la intersección Av. Antonio José de Sucre y Oe 10 perteneciente a la parroquia Centro Histórico, con un 90%, la segunda se registra en la intersección Av. de los Shyris y Naciones Unidas, perteneciente a la parroquia Ñaquito. Los resultados son coherentes pues en la parroquia centro Histórico están ubicadas gran cantidad de instituciones públicas que han establecido su horario de salida en el horario de 17:00 pm a 18:00 pm, de la misma manera en la parroquia Ñaquito están ubicadas la mayoría de instituciones privadas las cuales han establecido su horario de salida en el horario de 19:00 pm a 20:00 pm. De manera análoga podemos mencionar que las horas que menor cobertura registran es 10:00 am a 11:00 am en la intersección Av. 10 de Agosto y Av. Atahualpa, la razón obedece a que este

Herrera, Gómez.

Eficacia de la red social Twitter en la detección de casos de congestión vehicular

horario, es un horario en el cual la mayor parte de la población está dedicada a sus actividades de oficina, pues la intersección está en un sector donde hay empresas privadas.

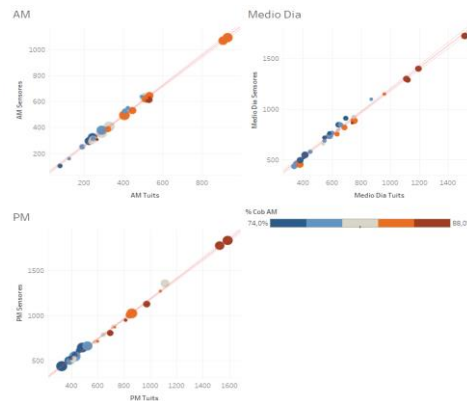


Figura 7. Situaciones de congestión detectadas por sensores y tuits por intersección y hora
Fuente: Elaboración propia

Conclusiones

El presente trabajo ha tenido como objetivo demostrar la eficacia del uso de sensores sociales como es el caso de la red social Twitter como herramienta de análisis y búsqueda del mejoramiento del tráfico en la ciudad y por tanto no es la que va a solucionar el tráfico mismo de la ciudad, sin embargo, corresponde a una herramienta de gran valor que puede ser parte de las soluciones, pues constituye una forma rápida, viable y de bajo costo para la toma de decisiones sobre congestión vehicular.

Utilizando la correlación entre la congestión del tráfico y la reacción social en Twitter, se establecen modelos que nos permiten estimar los atascos de tráfico utilizando sensores sociales.

En términos generales podemos mencionar que el porcentaje de cobertura de los tuits tanto por parroquias como por intersecciones es de al menos un 75%, lo cual avala que el estudio realizado cubre en un alto porcentaje los sectores en los cuales están instalados los sensores de flujo vehicular, ayudando así a que el proceso sea más preciso y lo más importante se pueda determinar en tiempo real.

De forma general se concluye que porcentaje de cobertura de los tuits en lugares donde no existen sensores de volumen de tráfico instalados, tanto por parroquias como por intersecciones es significativo (más del 90%), lo cual avala que el estudio realizado cubre en un alto porcentaje los sectores en los cuales no están instalados los sensores de flujo vehicular.

Referencias bibliográficas

- Agarwal, A. R. (2011). *Sentiment Analysis of Twitter data*. ACM Digital Library, 30-38.
- Arcila-Calderón, C., Ortega-Mohedano, F., Jiménez-Amores, J., & Trullenque, S. (2017). *Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático*. *Comunicacion Política*. doi:<https://doi.org/10.3145/epi.2017.sep.18>

- Avalos, H., & Gomez, E. O.-C. (2019). *Where to park? Architecture and implementation of an empty parking lot, automatic recognition system* . *Enfoque*.
- Delgado, R. (04 de Julio de 2016). *Memeburn*. Obtenido de Memeburn: <http://memeburn.com/2016/04/heres-how-big-data-is-being-used-to-fix-traffic-congestion/>.
- Digi. (01 de 08 de 2022). *Digi*. Obtenido de Digi: <https://es.digi.com/blog/post/smart-city-traffic-management-solutions>
- Feldman, R. y. (2007). *Mining Handbook. Advanced Approaches in Analyzing Unstructured Data*. New York, Estados Unidos : Cambridge.
- Hasan, A. y. (2018). *Machine learning-Based Sentiment Analysis for Twitter Accouts*. MDPI, 1-15.
- Herrera, S. (2016). *Analysis of the Factors Generating Vehicular Traffic*. *IEEE*, 133-137.
- Houhton, R. (2015). *Trasporte inteligete: Còmo mejorar la movilidad en la ciudad*. *Estadua Unidos*.
- IESE. (2017). *IESEinsight. España* .
- Matuschka, C. (2021). *Análisis de sentimientos usando la red social Twitter ¿qué sintieron los turistas que volaron en 2020 con seleccionadas aerolíneas sudamericanas?* *Revista de Turismo e Identidad*, 13(1), 55-71. doi:Matuschka
- Owen, R. y. (2012). *Mahout in action* . Manning Publications Co.
- Pozo, E. (2016). *Quito Camino a una movilidad sostenible* . Quito: Editorial General GK.
- Quito, A. (01 de Enero de 2011). *Distrito Metropolitano de Quito. la congestion vehicular sigue sin solucion en Quito*.
- Quito, M. d. (2014). *Diagnostico de la movilidad en el distrito Metropolitano de Quito para el Plan Metropolitano de desarrollo territoial PMOT*. Quito.
- Sankalp , S., Dheeraj , S., & Maniram, A. (2017). *Sentiment analysis based on text and emoticons*. *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*. India. Obtenido de <https://ieeexplore.ieee.org/xpl/conhome/7966475/proceeding>
- Schmind, H. (18 de 04 de 2018). *Tree Trager.dvi*. Obtenido de <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>.
- Varghese Babu, N., & Fabeela, A. (2021). *Multiclass Sentiment Analysis in Text and Emoticons of Twitter Data: A Review*. *Transactions on Computational Science and Computational Intelligence* , 61-68.
- Wang, L. (2015). *Classifications of clinical teeets using Apache Mahout*. *Classifications of clinical teeets using Apache Mahout*. Kansas City, Wuhan, China
- Zafarani, R. y. (2014). *Social Media Mining: An Introduction*. New York: Universityt Cambridge.