

N°8 NOTAS DE POLÍTICA

INDICADORES DE CALIDAD DE LOS SERVICIOS BÁSICOS: UN ENFOQUE DESDE TWITTER¹

Yoselin Arvelaiz², Daniel Barráez³ y Manuel Solorzano²

Resumen

Esta nota propone una metodología para construir indicadores de calidad de los servicios básicos a partir de Twitter. La metodología permite identificar fallas en diversos servicios básicos de uso doméstico (energía eléctrica, agua, transporte público, gas, recolección de residuos sólidos y combustible para vehículos), el tipo de fallas (apagón, cortes y desabastecimientos) y su ubicación por municipio o estado donde ocurrió la interrupción. Los índices construidos se pueden utilizar como proxies de los indicadores de los ODS. La metodología utiliza tweets que reportan escasez o fallas de servicios públicos residenciales en Venezuela. La bondad de ajuste de los algoritmos de localización fue del 95%. La metodología es una alternativa menos costosa a las encuestas estadísticas tradicionales, se puede implementar en tiempo real y se puede aplicar a cualquier otro país. La ubicación por municipio y estado proporciona información relevante para que los hacedores de política diseñen e implementen planes de infraestructura para el desarrollo, así como para identificar poblaciones y localidades vulnerables con el fin de implementar planes humanitarios y otras intervenciones. Los indicadores de fallas construidos pueden ayudar a llenar el vacío en las escasas estadísticas de fallas del servicio y las percepciones de calidad en muchos países.



¹ La versión original de este trabajo está disponible en tinyurl.com/r5572pkw.

² Universidad Central de Venezuela (yoselin.arvelaiz7@gmail.com, manuel.solorzano@ucv.ve).

³ Oficina de País de PNUD Venezuela (daniel.barraez@undp.org).

RESULTADOS PRINCIPALES

- Se propone una metodología para construir indicadores de calidad de los servicios básicos a partir de Twitter. La metodología se implementó para construir índices de calidad para los siguientes servicios: agua, energía eléctrica, gas doméstico, transporte público y recolección de residuos, así como combustible para vehículos.
- Los indicadores construidos dan cuenta de diferentes tipos de interrupciones (apagones, cortes y escasez), así como quejas de servicio, y están localizados a nivel del municipio o del estado donde ocurrió la interrupción. La bondad de ajuste de los algoritmos de localización fue del 95%.
- Los índices de agua, energía eléctrica, transporte público y recolección de residuos son proxies de los indicadores ODS 6.2.1, 7.1.1, 11.2.1 y 11.6.1, respectivamente. La agregación de todos los indicadores de servicios básicos es una variable proxy del indicador 1.4.1 de los ODS "Proporción de la población que vive en hogares con acceso a servicios básicos".
- La metodología es una alternativa menos costosa que las encuestas estadísticas tradicionales y se puede implementar en tiempo real.

RECOMENDACIONES DE POLÍTICA

- Los indicadores de fallas construidos pueden ayudar a llenar el vacío en las estadísticas de fallas de servicios básicos en muchos países y ayudar a los hacedores de política a dar seguimiento a las percepciones de la calidad de los servicios básicos.
- La ubicación por municipio y estado proporciona información relevante para que los hacedores de política diseñen e implementen planes de infraestructura para el desarrollo e identifiquen a la población vulnerable y las localidades para implementar planes humanitarios y otras intervenciones.

Introducción

La calidad de los servicios básicos tiene un impacto claro en la calidad de vida de la población y en el nivel de desarrollo de un país. Los servicios básicos son transversales a la mayoría de las metas de los ODS, y los ODS 1, 6, 7 y 11 enfatizan la importancia de la prestación de servicios básicos como condición para el desarrollo. Es común utilizar medidas de prestación de servicios básicos, como electricidad y agua potable, como proxies de la pobreza multidimensional, y está ampliamente aceptado que una provisión de calidad es necesaria para una buena gobernanza. El monitoreo de estos servicios proporciona información relevante para que los hacedores de políticas diseñen e implementen planes de infraestructura de servicios para el desarrollo. Además, las desigualdades en la provisión de estos servicios no afectan a toda la población de la misma manera, estas tienen un mayor impacto en las poblaciones vulnerables, por lo que monitorear la calidad de los servicios básicos es útil para identificar el grado de vulnerabilidad dentro de poblaciones y localidades específicas para implementar planes humanitarios y otro tipo de acciones.

Este documento propone una metodología para construir indicadores de la calidad de los servicios básicos utilizando tweets que informan sobre escasez, fallas o quejas para los siguientes seis servicios básicos: agua, energía eléctrica, gas doméstico, gasolina, recolección de desechos sólidos y transporte público. Este enfoque de deficiencia o insatisfacción está justificado. Las personas son más propensas a expresar insatisfacción que satisfacción, especialmente en presencia de interrupciones. Las fallas y la escasez no son inusuales en los países en desarrollo y están estrechamente relacionadas con las deficiencias de calidad. El enfoque a partir de Twitter también está justificado. Si bien los usuarios de Twitter no son una muestra representativa de la población, muchos actores relevantes utilizan esta red social para expresarse sobre temas de interés público: jefes y altos funcionarios de gobierno, parlamentarios, líderes políticos y sociales, ONG, medios de comunicación, influencers y ciudadanos de general. Estos actores son fundamentales para dar forma a las percepciones de las personas sobre muchos temas, como la calidad de los servicios básicos y sus opciones de mejora. Hoy en día, no cabe duda acerca de la importancia de las redes sociales en la discusión de los asuntos públicos. Los indicadores propuestos tienen algunas ventajas sobre las encuestas estadísticas tradicionales a la hora de medir la satisfacción de los servicios básicos. Es más rápido y menos costoso analizar tweets que recopilar encuestas y posteriormente analizar sus resultados. Además, los indicadores basados en Twitter pueden ser procesados en tiempo real. Por otro lado, los indicadores de fallas de los servicios o satisfacción construidos podrían ayudar a llenar el vacío en las estadísticas en muchos países. Aunque los proveedores de servicios registran habitualmente estadísticas de fallas, generalmente no están disponibles

para el público. Las estadísticas de satisfacción del usuario son más difíciles de encontrar porque requieren encuestas específicas de clientes que los proveedores u otras organizaciones no necesariamente realizan o, si lo hacen, no están disponibles para el público. Una ventaja adicional de Twitter sobre las encuestas es que el código de computador para procesar los tweets se puede compartir libremente, como se hace en este trabajo.

Recientemente, documentos, informes y artículos han utilizado con éxito el análisis de redes sociales para problemas de desarrollo. Solórzano (2018) utiliza Twitter para estudiar la pobreza multidimensional, y el informe PNUD (2021) emplea las redes sociales para analizar la información sobre el impacto de la contaminación en poblaciones vulnerables durante la pandemia de COVID-19. De acuerdo con estos análisis, empleamos una colección de herramientas de "minería de textos" (Aggarwal y Zhai, 2012; Python, 2021; The R Project for Statistical Computing, 2020; scikit-learn, 2021) para analizar los tweets. Nuestra metodología funciona en tres pasos. El primer paso es construir la base de datos de tweets que usaremos para identificar las fallas, utilizando las palabras clave para extraer los tweets relevantes, como apagones, cortes eléctricos o de agua. El siguiente paso es identificar los tweets que informan explícitamente de fallas o escasez. En el último paso, intentamos identificar el municipio o estado donde ocurrió la interrupción, principalmente analizando el texto de la publicación u otra información del usuario. Se respeta plenamente el anonimato u otra información privada; solo interesa la interrupción reportada y el lugar (municipio y estado) donde ocurrió y no los usuarios que reportan la falla. La metodología se implementó para los servicios básicos en Venezuela desde diciembre de 2019 hasta febrero de 2020. Para cada uno de estos servicios se construyó un índice y mapas de calor. La precisión de los algoritmos propuestos para identificar la ubicación de la falla resulta satisfactoria después de emplear pruebas estadísticas basadas en análisis de control de calidad.

Los principales hallazgos de este trabajo son los siguientes: analizando 1.671.869 tweets publicados en Venezuela del 01/12/2019 al 28/02/2020, se construyeron seis indicadores de calidad de los servicios básicos: agua, energía eléctrica, gas, combustible para vehículos, transporte público, y recolección de basuras. Los indicadores se localizaron a nivel de municipio y estado con un 95% de bondad de ajuste de los algoritmos de localización. Cuatro de estos indicadores son proxies de los indicadores de los ODS:

Cuadro 1. Indicadores de los ODS

Proxy propuesto	Indicador ODS
Agua	6.2.1 Proporción de la población que utiliza (a) servicios de saneamiento gestionados de forma segura y (b) una instalación para lavarse las manos con agua y jabón.
Energía eléctrica	7.1.1 Proporción de población con acceso a electricidad.
Transporte público	11.2.1 Proporción de la población que tiene acceso conveniente al transporte público, por sexo, edad y personas en condición de discapacidad.
Recolección de basuras	11.6.1 Proporción de residuos sólidos urbanos recogidos y gestionados en instalaciones controladas sobre el total de residuos municipales generados, por ciudades.
Todos los indicadores agregados	14.1 Proporción de la población que vive en hogares con acceso a servicios básicos.

La metodología propuesta es una alternativa menos costosa a las encuestas estadísticas tradicionales, se puede implementar en tiempo real y los indicadores construidos pueden ayudar a llenar el vacío en las estadísticas de fallas del servicio o satisfacción en muchos países. Los algoritmos propuestos se basan en técnicas de minería de textos que son fácilmente modificables para reconocer otros listados de municipios e incluso idiomas, lo que hace que la metodología sea adaptable a otros contextos geográficos.

Este documento de trabajo contiene tres secciones adicionales. En la sección de metodología, presentamos los algoritmos para la detección de fallas y la ubicación de fallas regionales y damos algunos ejemplos para mostrar cómo funcionan. En la sección de resultados mostramos los índices, su evolución en el tiempo, los mapas de calor e identificamos los municipios y estados más afectados. Finalmente, proporcionamos algunas conclusiones.

Datos y metodología

En esta sección se presenta la metodología con la que intentamos identificar diferentes tipos de irregularidades en la prestación de servicios básicos que afectan la calidad de vida de la población. Las irregularidades varían según el tipo de servicio, como cortes de luz o agua, escasez de gasolina, escasez de gas que ocasiona la venta o distribución irregular de gas en cilindros o su recarga, dificultades en el transporte público, como escasez o ausencia de unidades de transporte. En lo que sigue, para simplificar el lenguaje, usaremos la palabra "falla" para referirnos a todos estos diferentes tipos de alteraciones en el servicio.

Construcción de las bases de datos. El primer paso en la metodología es construir la base de datos que contiene el conjunto de tweets "candidatos" para informar sobre las fallas o quejas de servicios básicos. Por tweet candidato se refiere a una publicación que probablemente informa de las fallas, pero no están seguros de la ocurrencia de la falla reportada. Este paso se logra utilizando las palabras claves de búsqueda en la API de Twitter relacionada con las fallas del servicio básico para extraer las publicaciones relevantes, como apagones, cortes eléctricos o de agua. La base de datos utilizada en este trabajo está compuesta por las 1.671.869 publicaciones del 01/12/2019 al 28/02/2020. Los tweets seleccionados se filtraron para considerar solo aquellos que se publicaron en Venezuela y para enfocarse en los siguientes seis servicios básicos: agua, electricidad, gas, gasolina, saneamiento y transporte público.

Detección de fallas. Después de un análisis de datos exploratorio de la base de datos, se construye un "diccionario" o colección de palabras para identificar fallas para cada servicio básico considerado, lo que resulta en seis diccionarios. En el Cuadro 1 se muestran algunos ejemplos de palabras contenidas en los diccionarios. El algoritmo de detección de fallas funciona de la siguiente manera. Para una publicación determinada, si contiene una combinación de palabras específica en uno de los diccionarios, se detecta una falla y el tweet se asigna a la base de datos del servicio de fallas correspondiente. Vale la pena señalar que el algoritmo de detección de fallas identifica una secuencia de palabras y no palabras aisladas. Por ejemplo, si la publicación de texto es "sin agua en Maracaibo", el algoritmo reconoce la publicación como una falla por la expresión "sin agua en". Sin embargo, si el texto es "sin dinero no se puede comprar un vaso de agua", el algoritmo no identifica una falla, aunque las palabras "sin" y "agua" están en el texto. Como un tweet puede expresar múltiples fallas en los servicios (por ejemplo, "sin agua ni gasolina"), cada publicación se compara con los seis diccionarios y puede incluirse en más de una base de datos de fallas. Por supuesto, algunas publicaciones permanecerán excluidas de las seis bases de datos de fallas. Al finalizar el proceso, tenemos seis nuevas bases de datos con las fallas detectada en los tweets procesados. Estas bases de datos también almacenan el tiempo de publicación de los tweets para seguir la evolución de la falla a lo largo del tiempo. Cada base de datos de fallas proporciona la información de fallas a nivel nacional.

Cuadro 2. Diccionario de palabras clave para las diferentes categorías

Electricidad	<ul style="list-style-type: none"> ▪ Apagón ▪ Apagones ▪ Luz ▪ Electricidad ▪ Eléctrico (s) ▪ Corpoelec
Agua	<ul style="list-style-type: none"> ▪ Sin agua ▪ Hidrocapital ▪ Hay una falla ▪ Cortes de agua
Transporte	<ul style="list-style-type: none"> ▪ Metro ▪ Furgoneta (s) ▪ Carro(s) ▪ Pasajero (s) ▪ Metrobus ▪ Autobús ▪ Transporte (s) ▪ Motocicleta (s) ▪ Automóvil (s) ▪ Vehículo (s) ▪ Camión (s)
Gas	<ul style="list-style-type: none"> ▪ Cilindro ▪ Gas ▪ Leña ▪ Cocina ▪ Estufas ▪ Carbón ▪ Carbones
Recolección de basuras	<ul style="list-style-type: none"> ▪ Basura (s) ▪ Colección ▪ Inodoro (s) ▪ Alcantarillado (s) ▪ Saneamiento (s) ▪ Desperdicio (s) ▪ Excreta (s) ▪ Chatarra (s) ▪ Residuo (s) ▪ Contenedor (s)
Gasolina	<ul style="list-style-type: none"> ▪ Gasolina ▪ Combustible ▪ Estación de servicio

Identificación del municipio/estado. Luego, intentamos identificar el municipio y el estado. El algoritmo explora tres campos de información en el tweet, que están jerarquizados en el siguiente orden. Primero, el texto del tweet puede proporcionar información clara sobre la ubicación de la falla. En segundo lugar, la ubicación de transmisión del tweet: la información de ubicación que emite el dispositivo del usuario al generar el tweet. Esta información puede incluir el país, estado o ciudad del usuario. En tercer lugar, la ubicación de la ciudad, municipio o estado del usuario en el momento de crear su cuenta.

Las dos últimas fuentes de información solo están disponibles si el usuario las autoriza. El algoritmo no utiliza identificación personal; se respeta el anonimato en la información del usuario. Una lista de municipios y estados, como un diccionario de municipios, se construye de manera similar a los diccionarios de fallas antes mencionados.

Se realiza una prueba estadística para verificar la bondad de ajuste de los algoritmos utilizados para detectar fallas y sus ubicaciones. Para esta prueba se seleccionaron al azar 300 tweets con su detección de fallas y los resultados de la ubicación, seguidos de una verificación humana de la precisión de los resultados. La identificación y la localización de la falla fueron correctas el 95% de las veces. Todos los requisitos estadísticos se cumplieron (Gutiérrez y De la Vara, 2013).

Finalmente, construimos el índice de falla o calidad por agregación a nivel nacional y regional con sus respectivos mapas de calor.

Los siguientes ejemplos ilustran cómo funciona el procedimiento. Ejemplo A "sin electricidad en Maracaibo", ejemplo B "sin electricidad en casa" y ejemplo C "cuando escucho estos argumentos siento como una descarga eléctrica en mi cabeza".

1. Construcción de bases de datos. Los tres ejemplos se incluyen en la base de datos debido a la palabra "electricidad" en el texto de cada tweet.

2. Detección de fallas. El ejemplo A se identifica como una falla por la expresión "sin electricidad". El ejemplo B también se identifica como una falla por la expresión "no hay electricidad en". El ejemplo C no se identifica como una falla porque no hay una expresión en los diccionarios de fallas eléctricas que coincida con una falla. La palabra aislada "electricidad" no es suficiente para identificar una falla. Por tanto, el ejemplo C queda excluido del proceso.

3. Identificación del municipio/estado. En el ejemplo A, con la expresión "en Maracaibo" identificamos el municipio de Maracaibo en el estado Zulia de la lista de municipios y estados. En el ejemplo B, dado que se identifica la expresión sobre el lugar, tratamos de identificar la ubicación de transmisión del tweet o la ubicación en el momento de crear la cuenta de usuario.

Resultados

Las seis bases de datos de fallas de servicio que contienen 515.439 tweets proporcionan toda la información necesaria para construir los índices de fallas. El Cuadro 3 muestra la distribución de tweets por servicio.

Resultados de fallas

Cuadro 3. Distribución de tweets por detección de fallas

	Electricidad	Agua	Transporte	Gasolina	Gas	Recolección de basuras
Total de tweets	212,847	125,259	63,921	62,433	26,290	24,689
% participación	41	24	13	12	5	5

Cada índice se normaliza al promedio diario durante los tres meses de estudio; un valor de 100 corresponde a la media muestral. En las siguientes figuras mostramos el promedio móvil de siete días (Brockwell & Davis, 1996) a nivel nacional.

Figura 1. Índice de agua

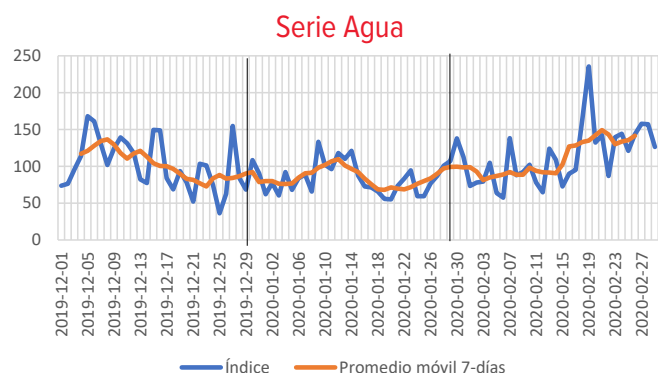
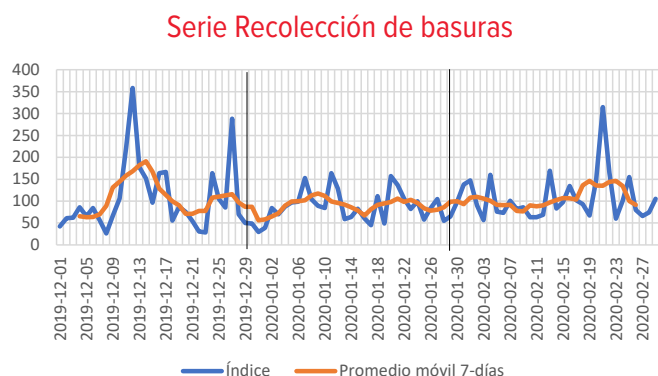


Figura 2. Índice de saneamiento



El rango de 36 a 236 para el índice de agua en la figura 1 se explica por una combinación de los efectos de la temporada de lluvias (muchos valores por debajo de 100) y cortes específicos del servicio de agua (valores por encima de 100). Para el índice de saneamiento en la figura 2, los informes muestran baja intensidad, excepto los días 12-diciembre-2019, 27-diciembre-2019 y 21-enero-2020. La falla de promedio móvil reporta un incremento durante diciembre, uno de los meses con mayor actividad comercial.

Figura 3. Índice de combustible para automóviles

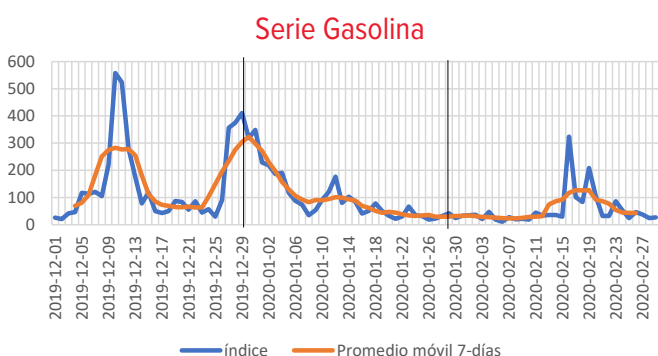
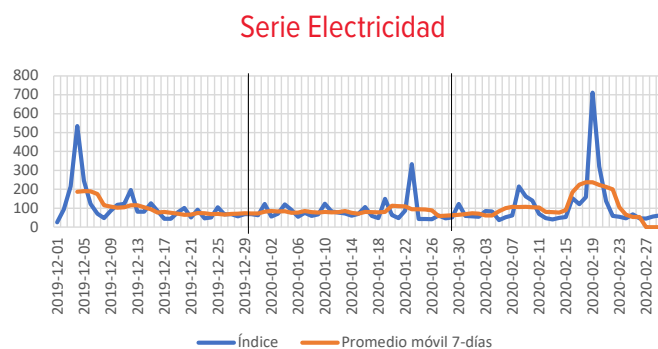


Figura 4. Índice de energía eléctrica



En la Figura 3, el índice identifica la escasez de gasolina de mediados de diciembre de 2019. Se observaron largas colas en las estaciones de servicio. Durante enero de 2020, el suministro de combustible se normalizó parcialmente. A mediados de febrero de 2020, volvió a aparecer el desabastecimiento, especialmente en los estados fronterizos de Táchira, Barinas y Mérida, generando muchas quejas en Twitter. El comportamiento del índice es consistente con las noticias en los medios.

Podemos ver en la figura 4 que este índice captura con éxito los apagones eléctricos. Los valores del índice por encima de 300 coinciden con las interrupciones informadas en al menos cinco estados. Tenga en cuenta que los índices de gasolina y electricidad son menos volátiles que los del agua y el saneamiento.

Figura 5. Índice de gas

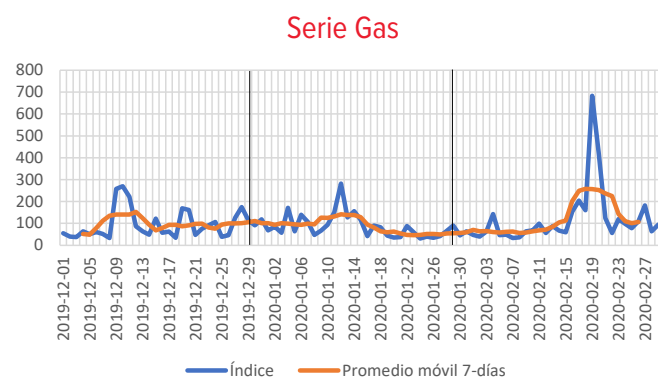
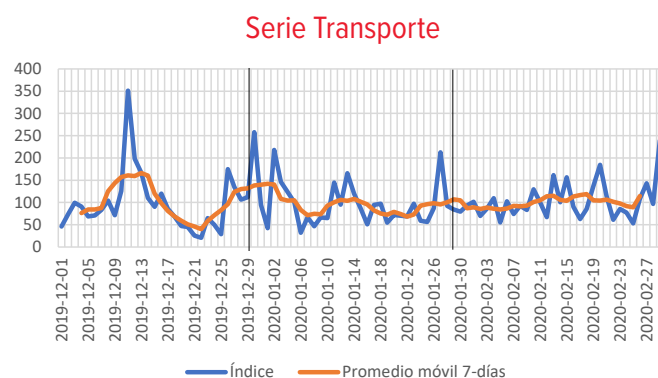


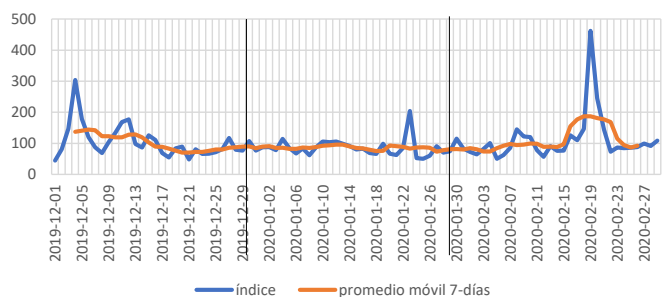
Figura 6. Índice de transporte público



La figura 6 muestra que el índice de servicios de transporte está altamente correlacionado con el índice de combustible automotriz en la figura 3 porque casi todas las unidades de transporte en el país usan combustible. Como se observa en los gráficos del 1 al 6, todos los índices tienen un comportamiento natural e intuitivo.

Figura 7. Índice de servicios básicos integrados

Serie Servicios básicos integrados



Para construir un indicador proxy del indicador 1.4.1 de los ODS "Proporción de la población que vive en hogares con acceso a servicios básicos", construimos un índice con la agregación de todas las bases de datos de fallas de los servicios (excluyendo el indicador de gasolina), el índice consolidado de los servicios básicos se muestra en la figura 7. La exclusión de la escasez de gasolina se justifica porque este servicio no está incluido en los metadatos del indicador 1.4.1 ODS (UNSTATS).

Resultados de localización

El algoritmo de localización pudo identificar el estado y el municipio del 71% y 64% de los tweets respectivamente. Estos resultados nos permiten identificar las áreas con mayores dificultades, que se muestran en los siguientes mapas de calor en las figuras 8 y 9.

Figura 8. Número de fallas reportadas en servicios básicos en Venezuela a nivel estatal

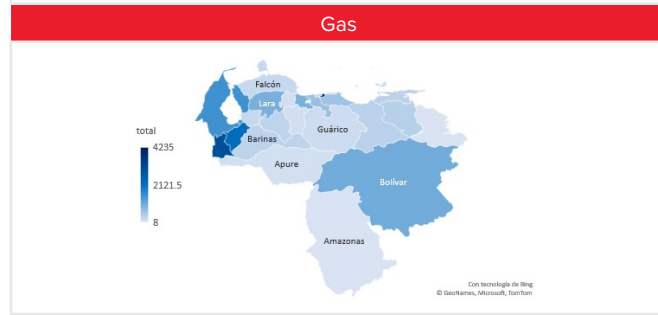
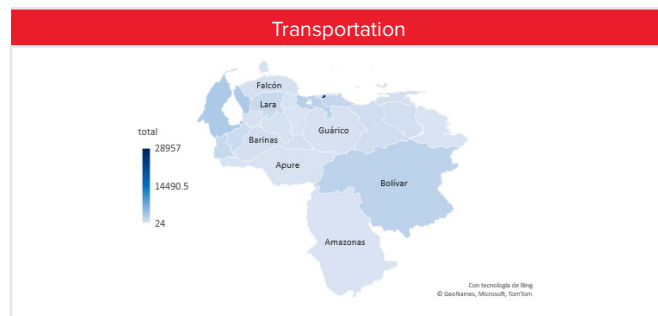
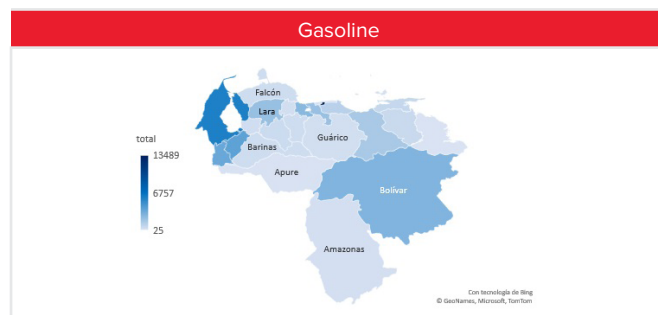
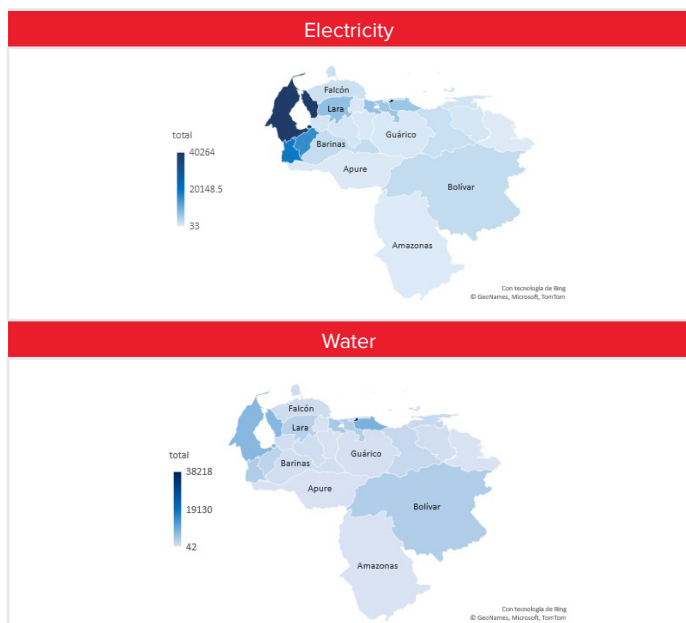
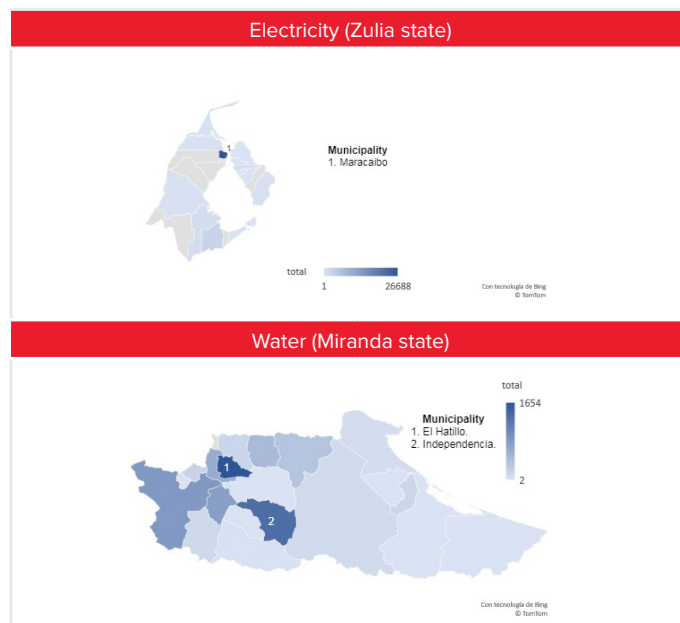
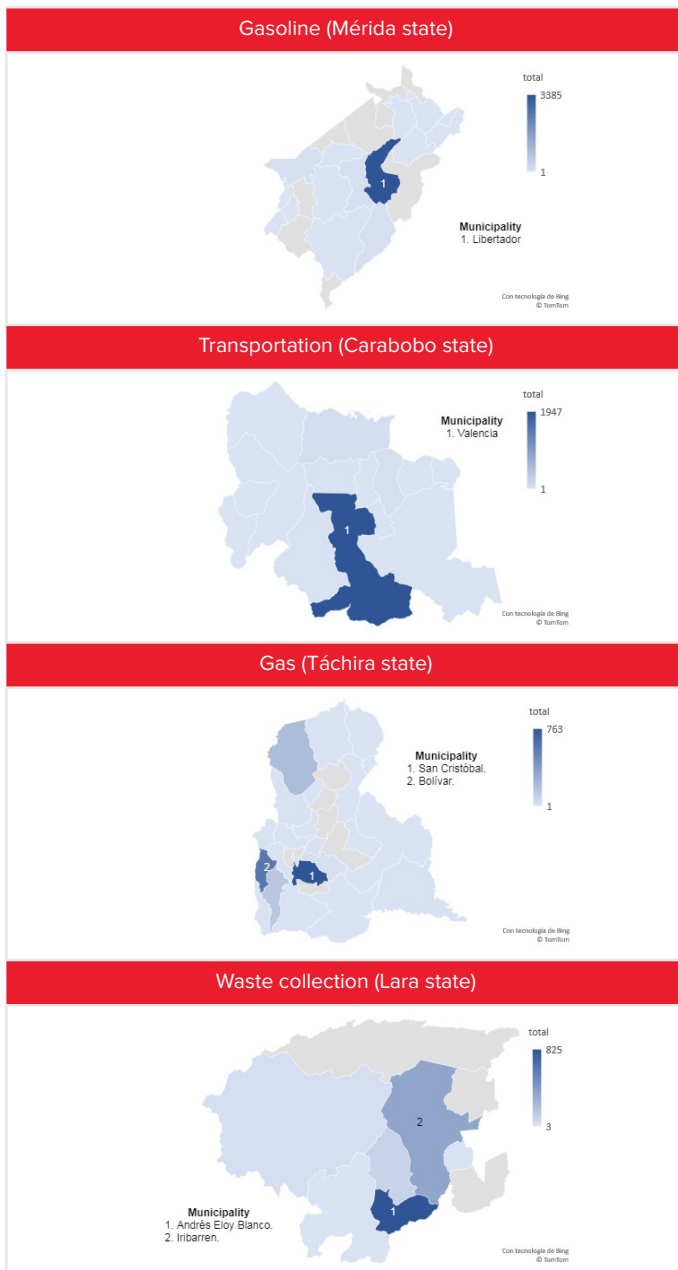


Figura 9. Los municipios con mayor nivel de fallas





Vale la pena señalar que la detección de fallas no es una función simple de la población o los usuarios de Internet en cada estado; tenemos una baja correlación entre ellos, como podemos ver en el Cuadro 4.

Cuadro 4. Detección de fallas por servicio y población o correlación de usuarios de Internet

	Electricity	Water	Gasoline	Public transport	Gas	Waste collection
Poblacion	0.64	0.41	0.5	0.28	0.4	0.39
Internet	0.57	0.62	0.56	0.49	0.48	0.57

Por otro lado, aunque no tenemos datos reales de fallas para validar nuestros índices, la falta de dichos datos es una de las motivaciones de esta metodología, las interrupciones

del servicio público reportadas por los usuarios de Twitter son ampliamente confirmadas por reportes de medios y tweets informativos de las instituciones proveedoras de servicios.

Conclusiones

Los resultados muestran que la metodología propuesta es capaz de identificar fallas en diversos servicios (energía eléctrica, agua potable, transporte, gas doméstico, recolección de desechos sólidos y combustible de vehículos), el tipo de fallas (apagón, cortes y desabastecimiento) y la ubicación por municipio o estado donde ocurrió la perturbación. El algoritmo de localización también identifica los municipios y estados más vulnerables. Esta identificación es una información valiosa para la toma de decisiones sobre priorización de zonas que necesitan inversiones en infraestructura de servicios y planes de respuesta humanitaria. La metodología es una alternativa menos costosa a las encuestas estadísticas tradicionales y se puede implementar en tiempo real. Los diccionarios de fallas y localización empleados por los algoritmos son fácilmente modificables e incluso pueden traducirse a otros idiomas, por lo que son adaptables a otros países. Los indicadores de fallas construidos pueden ayudar a llenar el vacío en las estadísticas de fallas del servicio en muchos países.

Referencias

Aggarwal, C., & Zhai, C. (2012). Mining Text Data. Springer Science+Business Media.

Brockwell, P., & Davis, R. (1996). Introduction to Time Series and Forecasting. New York: Springer-Verlag, New York Inc.

Gutiérrez, H., & De la Vara, R. (2013). Control Estadístico de la calidad y Seis Sigma (Third ed.). México: McGraw-Hill/Interamericana Editores, S.A. DE C.V.

Python. (2021). re — Regular expression. Retrieved from Python Documentation: docs.python.org/3/library/re.html

scikit-learn. (2021). scikit-learn. Retrieved from Machine Learning in Python: scikit-learn.org/stable

Solorzano, M. (2018). Las voces de Twitter sobre las dimensiones de pobreza en Venezuela: Un enfoque de Big Data. Caracas: Central University of Venezuela.

The R Project for Statistical Computing. (2020, November 18). Text Mining in R. Retrieved from Text Mining Package: cran.r-project.org/web/packages/tm/tm.pdf

UNDP. (2021). Narratives and Information Pollution on Vulnerable Groups During The COVID-19 Pandemic: Panama and Venezuela cases. Retrieved from UNDP: www.ve.undp.org/content/venezuela/es/home/library/human_development/narrativas-y-contaminacion-informativa-sobre-grupos-vulnerables-htm

United Nations. (n.d.). Transforming our world: the 2030 Agenda for Sustainable Development. Retrieved from UN, Department of Economic and Social Affairs: sdgs.un.org/2030agenda

UNSTATS. (n.d.). SDG Indicators. Retrieved from Sustainable Development Goals: unstats.un.org/sdgs/indicators/indicators-list

UNSTATS. (n.d.). SDG Indicators Metadata repository. Retrieved from Sustainable Development Goal indicators: unstats.un.org/sdgs/metadata



www.latinamerica.undp.org



twitter.com/PNUDLAC



www.facebook.com/pnudlac



www.instagram.com/pnudlac



www.linkedin.com/company/pnudlac



www.youtube.com/PNUDLAC