

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



El futuro es de todos

DNP
Departamento
Nacional de Planeación

ANÁLISIS DE CONTENIDO COMPARTIDO SOBRE EL PND EN TWITTER

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Subdirección General Sectorial.

Sector

Planeación

Lenguaje

R y Python.

Fuente de datos

Twitter.

Presentación

En el marco de las negociaciones del Plan Nacional de Desarrollo, es muy importante conocer las opiniones divulgadas por grupos de interés específicos para tomar decisiones estratégicas que faciliten la negociación. La aplicación que aquí se presenta, desarrollada por la Unidad de Científicos de Datos, permite analizar tweets referentes a algunos sectores y en ventanas de tiempo específicas. Los resultados evidencian palabras, sentimiento, alcance y temas de interés que caracterizan los *tweets*, al igual que facilita la identificación de actores con interés por cada sector. Esta aplicación está dirigida a tomadores de decisiones estratégicas y al grupo de comunicaciones del Departamento Nacional de Planeación.

Within the framework of the negotiations of the National Development Plan, it is very important to know the opinions shared by specific interest groups in order to take strategic decisions that facilitate the negotiation. The application presented here, developed by the Data Scientists Team, makes it possible to analyze tweets referring to some sectors and in specific time windows. The results show words, feelings, impact and topics of interest that characterize the tweets and facilitate the identification of actors with interest in each sector. This application is aimed at strategic decision-makers and the communications group of the National Planning Department.

Objetivo general

Analizar los tweets compartidos que hacen referencia al PND y a sectores estratégicos.

Objetivos específicos

1. Crear una rutina de descarga y actualización de *tweets* sobre los temas de interés.
2. Desarrollar algoritmos de procesamiento y análisis de texto para identificar tweets referidos al DNP y a sectores estratégicos.
3. Desarrollar una aplicación interactiva que permita identificar los actores principales, los *tweets* más importantes, los sentimientos y los términos más característicos.

Metodología

El procedimiento desarrollado para el análisis puede resumirse en cinco etapas: (1) la descarga de la información de Twitter, (2) el procesamiento del texto de los *tweets*, (3) la identificación de *tweets* de interés a partir de palabras clave, (4) la construcción de un modelo para evaluar el sentimiento de los tweets y (5) la construcción de herramientas para visualizar los resultados. Para la primera etapa, se construyó un código en Python que utiliza un API (*Application Programming Interface*) de Twitter para descargar los Tweets de interés. Este código permite organizar los Tweets de numerosos usuarios junto con su fecha y hora de publicación.



El futuro es de todos

DNP
Departamento
Nacional de Planeación

Para realizar el análisis de texto en la segunda etapa, se somete el contenido de los tweets a un proceso de limpieza en el que se eliminan números, signos de puntuación, espacios dobles y caracteres no imprimibles, tras lo cual se pasa el texto a minúsculas y se eliminan las palabras que no agregan significado al texto (artículos, preposiciones, pronombres, etc.). En la tercera etapa, se utilizan las palabras clave definidas por las distintas dependencias del DNP con base en el juicio de los expertos de los siguientes sectores:

- TIC
- Turismo y economía naranja
- Social
- Minería e hidrocarburos
- Agrícola
- Justicia y seguridad
- Ciencia, tecnología e innovación
- Transporte
- Renovación urbana
- Energía
- Mejora regulatoria
- Desarrollo productivo
- Servicios públicos

Así, se creó un algoritmo para identificar estas palabras en los *tweets* y etiquetarlos según hicieran referencia a uno o a varios sectores. Esta misma metodología se utilizó para identificar los *tweets* que mencionan temas relacionados con el DNP y con el Plan Nacional de Desarrollo. Una representación esquemática del procedimiento que realiza el algoritmo se presenta en la Figura 1.

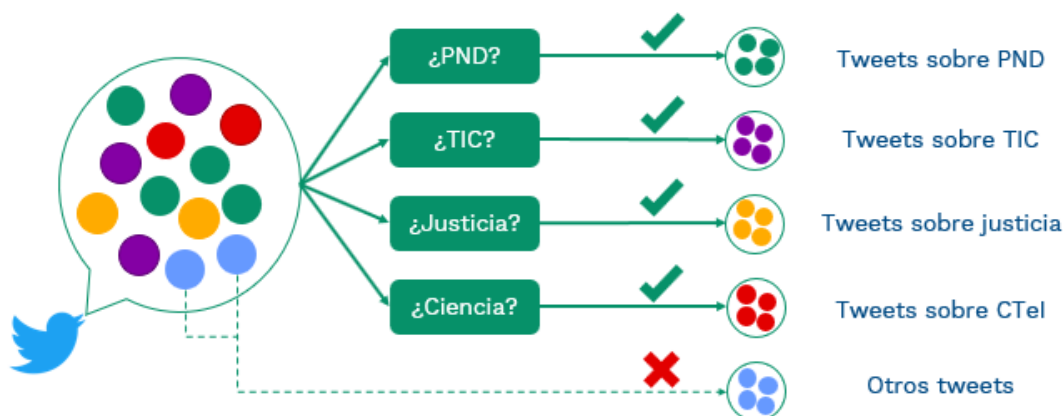


Figura 1: Esquema de la clasificación que realiza el algoritmo. Algunos grupos pueden superponerse, pues un mismo tweet puede tener varias etiquetas.

Adicionalmente, para identificar el sentimiento de los tweets (positivos o negativos) se entrenó un modelo de clasificación. En primera instancia, se utilizaron palabras positivas y negativas, así como emoticones utilizados por los usuarios para construir una base de entrenamiento de tweets con sentimiento positivo y negativo. Esta base fue utilizada para entrenar una red neuronal consistente en una capa de *embedding* (vectorización), una capa de *Gated Recurrent Units* (GRU) y dos capas completamente conectadas. Este modelo se utiliza para cuantificar (en una escala de 0 a 1) el sentimiento de cada tweet y calificarlo como positivo, negativo o neutro.

Finalmente, se creó una aplicación utilizando el paquete *shiny* de R en la cual se pueden filtrar los *tweets* por fecha, sector, relación con el PND, sentimiento (positivo, neutro, negativo) e incluso por palabras clave definidas por el usuario. Aplicando los filtros de interés, podrán visualizarse después los usuarios con mayor interés por cada sector, los que tienen mayor alcance (medido en número de *retweets*), los *tweets* más recientes, los más positivos o negativos, los de mayor alcance (sea por número de *likes* o por número de *retweets*), y las palabras más usadas en forma de nube de palabras. En la figura 2 se resaltan algunas de las propiedades más características de la aplicación. Esta aplicación se ejecuta desde un servidor local donde se programó la



El futuro es de todos

DNP
Departamento
Nacional de Planeación

ejecución automática de los algoritmos de descarga y procesamiento de *tweets* para mantener el contenido actualizado diariamente. Para su divulgación, se desarrollaron un manual de usuario y dos tutoriales en video.

Alcance	Sentimiento	Búsqueda personalizada
 Likes  Retweets	 Positivo  Neutro  Negativo	Buscar tweets que contengan... <input type="text" value="Presidente"/> Y <input type="text" value="Colombia"/>

Figura 2: Características que se analizan en la aplicación desarrollada.

Resultados

Desde la aplicación desarrollada, se analizaron más de 533.000 tweets con corte a marzo 18 de 2019. Allí, se segmentó el análisis para identificar los usuarios con mayor interés e influencia en cada sector y los temas sobre los cuales hablan. Un ejemplo de estos resultados se muestra en la figura 3, donde puede observarse, por ejemplo, cómo el tema del glifosato es de los más mencionados respecto a justicia y seguridad, permitiendo incluso vislumbrar etiquetas (*hashtags*) promovidas por los usuarios como la de #NoMásVeneno. También pueden identificarse *tweets* de gran interés, como los *tweets* negativos sobre el PND que tienen mayor alcance.



Figura 3: Análisis de tweets sobre el sector social



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación

Conclusiones

1. El análisis brinda información de gran importancia para tomar decisiones sobre la estrategia de comunicaciones del DNP.
2. El algoritmo permite identificar tweets sobre distintos sectores y el PND.
3. La aplicación permite al usuario realizar análisis por fechas, sectores y sentimientos.
4. El análisis realizado permite identificar las palabras que caracterizan los *tweets* de interés.

Socialización

La aplicación fue presentada por la Unidad de Científicos de Datos y por la Subdirección General Sectorial a las dependencias sectoriales del DNP y por la Dirección General a distintos actores de la rama ejecutiva.