

Taller Segundo Corte Minería de Datos

Donald Trump es un personaje controversial, así mismo es un usuario muy activo de twitter expresando sus ideas y estableciendo un contacto directo con los usuarios de Twitter.

De acuerdo a los tweets de Trump se debe realizar un desarrollo de minería de texto identificando las características semánticas:

- Realizar un análisis descriptivo del dataset proporcionado
 - Identificar y cuantificar el software o sistema operativo que fue publicado
 - visualizar dentro del rango de 24 horas como distribuyen los tweets, identificando el origen de tweet
- analizando las frecuencias de las palabras de forma global
 - Realizar un análisis de frecuencia basándose en los dispositivos IOS y Android
- Eliminar los stopwords encontrados en el corpus, realizar el análisis de frecuencias.
- usando el corpus de malas palabras cuantificar cuántos tweets y qué porcentaje representan estos tweets en donde Donald Trump usa malas palabras.
 - Realizar este analisis entre los dispositivos IOS y Android
- Desarrollar documento que evidencia el proceso desarrollado, analisis de los diferentes puntos, graficas y conclusiones del analisis realizado.

Para el desarrollo de este taller utilice todos los elementos y herramientas vistos en clase, y conceptos de análisis estadístico para el desarrollo del taller.

Nota: subir código fuente y documento

Fuente del Dataset: <http://www.trumptwitterarchive.com>