

Blog

Edward R. Rivera Rivera

Marzo 2021

¿Midiendo Cuantitativamente El Moldear La Opinión Pública?

Sin lugar a duda, y lo queramos o no, los medios de comunicación tienen una influencia incalculable en determinar la llamada opinión pública. Lo podemos observar a diario: en nuestras interacciones con amigos, familiares y en las redes sociales. También hemos vistos acusaciones de parte de políticos y sus “colaboradores cibernéticos no oficiales” (fotutos) que ciertos periodistas intentan moldear la opinión pública en base de alguna agenda peculiar escondida. Como persona curiosa siempre me he preguntado cuan ciertas son estas acusaciones y, de ser ciertas, como científico me pregunto que métricas podemos usar para determinar si en realidad se está intentando moldear la opinión pública. Pero ¿cómo se puede lograr ese objetivo?

Aunque no hay métrica alguna (a mi conocimiento) para poder medir científica y objetivamente si se está moldeando la opinión pública, si existen varios mecanismos para procesar textos como datos. Tradicionalmente, cuando una persona quiere analizar algún problema con algún método cuantitativo, el o ella recopilan datos numéricos y se realiza alguna prueba estadística y se llegan a conclusiones en base a los hallazgos. Para poder lograr un análisis cuantitativo para datos textuales, se toma un documento (o documentos) y se crea una lista de todas las palabras únicas en dicho documento y se cuenta la frecuencia de cada una de las palabras usadas—en vernáculo técnico, lo convertimos en lo que se conoce en inglés como “document-feature matrix (DFM)”. Tomemos por ejemplo las siguientes oraciones “Mi mama me ama. Mi mama me mima. Extraño a mi mama.”

```
## Document-feature matrix of: 1 document, 7 features (0.0% sparse).  
##           features  
## docs      mi mama me ama mima extraño a  
##  text1  3      3  2   1      1      1  1
```

Como podemos observar, el documento compuesto por las tres oraciones contiene las palabras “mi” y “mama” tres veces, la palabra “me” dos veces, etc. . . En este ejemplo, el proceso descrito puede realizarse fácilmente a mano ya que no tiene muchas palabras únicas; sin embargo, el proceso se convierte demasiado tedioso cuando los documentos son mas extensos. Por fortuna, avances en tecnología y computación han facilitado este proceso para poder efectuar análisis de texto como datos.

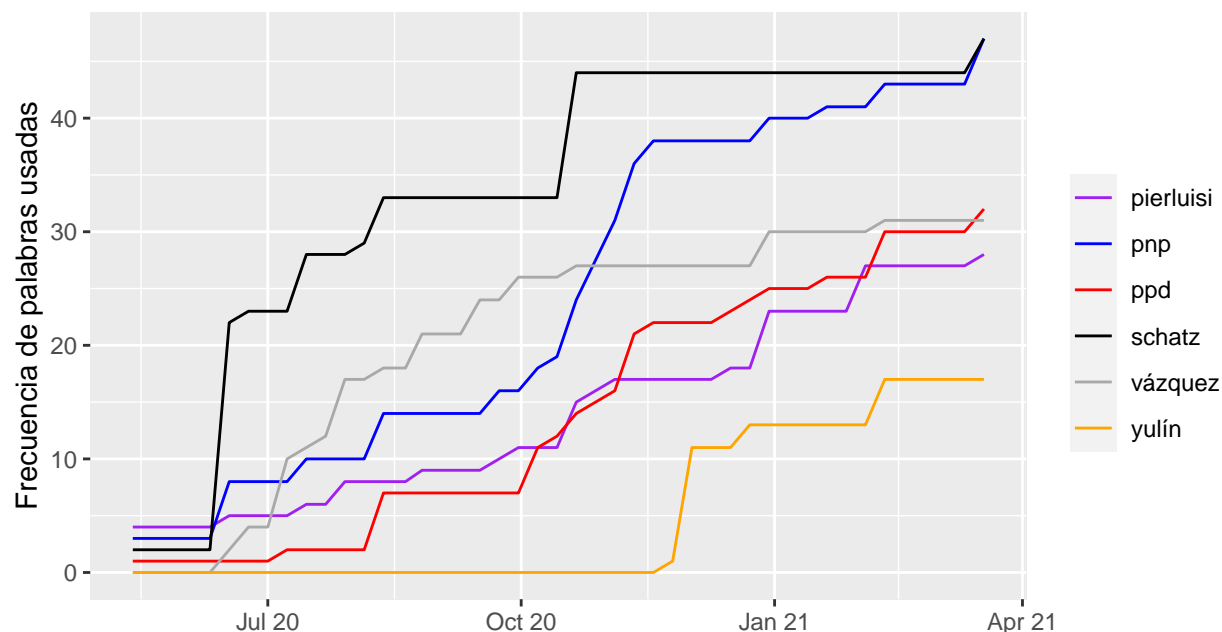
Ahora, para contestar la interrogante de medir cómo se moldea la opinión pública hay que deconstruir y dividir el problema en muchas secciones. Para comenzar mi proyecto, seleccioné al abogado y personalidad radial Jay Fonseca por varias razones. Primero, en los últimos años el Lcdo. Fonseca ha extendido su influencia tanto en la radio como en las redes sociales. Segundo, Jay escribe una columna semanal (Esto tiene salvación) en el periódico Primera Hora. Y último, muchos usuarios de las redes sociales están convencidos que Jay Fonseca moldea la opinión pública para cumplir con una agenda ulterior (para algunos, una agenda socialista y para otros, una agenda neo-liberal). Por esos motivos, escogí los escritos de Jay Fonseca como un buen punto de partida. Además, los datos son los datos.

Ya habiendo seleccionado las columnas de Jay Fonseca como punto de partida, comencé a crear mi propia base de datos recopilando todas las columnas disponibles en Primera Hora. Aunque es de mi conocimiento que Jay lleva años escribiendo para el periódico, Primera Hora solamente tiene disponible los artículos

publicados desde 13 de mayo del 2020. Luego de haber añadido todas las columnas a mi base de datos, procedí a convertir todos los documentos a DFM (por sus siglas en inglés). Para tener una idea mas clara, filtré e hice una representación gráfica de ciertas palabras

Esto Tiene Salvación: Los Datos de Jay Fonseca

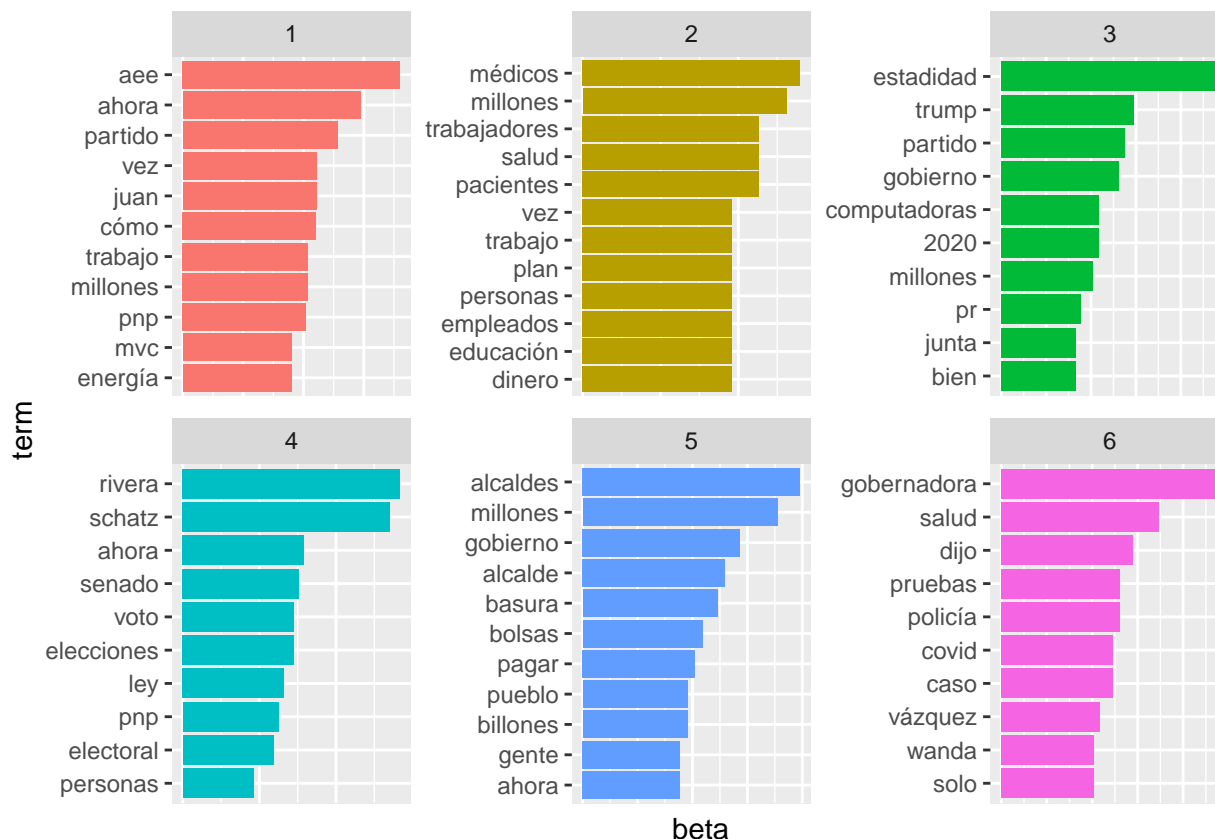
Breve representación gráfica de las columnas escritas en Primera Hora



Fuente: Primera Hora
Artículos compilados desde el 13 de mayo 2020
Palabras escogidas fueron seleccionadas arbitrariamente

Aunque la gráfica es una interesante y muestra información valiosa, yo seleccioné de manera arbitraria las palabras incluidas en el análisis y para llevar una medición mas justa y seria, no puedo usar esta gráfica para llegar a conclusiones. Una alternativa para usar este tipo de gráfica (ya que es útil para apreciar las tendencias al pasar del tiempo) es crear “diccionarios”. Un diccionario en análisis textual cuantitativo se define como una lista con palabras que se asocian a un tema o una idea en particular. Por ejemplo, palabras como “Rivera Schatz, palma, Pierluisi, etc...” pueden asociarse al Partido Nuevo Progresista; cuando se hayan creado los diferentes diccionarios, entonces podemos hacer un análisis mas específico para apreciar las tendencias de palabras asociadas a un tema o idea en un tiempo determinado. Esta tarea requiere de un tiempo mas extenso para poder crear certeramente diferentes diccionarios. Lo tengo en agenda, se las debo.

En lugar de una gráfica lineal, decidí usar una técnica de modelado de temas (topic modeling) llamada “Latent Dirichlet allocation” (LDA de manera mas simple). Este método nos permite tratar cada documento como una mezcla de temas y cada tema como una mezcla de palabras (Topic Modeling, 2021). Para evitar una explicación técnica y crear mas confusión, aquí les grafico los resultados del método LDA.



Como pueden apreciar, la gráfica está dividida en seis secciones diferentes que representan los temas. Dentro de esos temas, están las palabras que el algoritmo del modelo calculó pertenecían a dicho tema basado en la frecuencia con que se utilizaban las palabras. Por ejemplo en el tema número 1 vemos palabras como “partido”, “pnp” y “mvc” lo cual pudiera sugerir que uno de los temas que Jay Fonseca escribe es acerca de partidos políticos (no debe ser sorpresa para nadie que este familiarizado con el señor Fonseca). En el tema 2, vemos que el algoritmo regresó palabras como “médicos”, “salud”, “trabajadores”, “empleados” lo cual sugiere que Jay Fonseca escribe sobre diferentes agencias de gobierno. En el tema 3, palabras como “estadidad”, “trump”, “2020” pudiera sugerir que el tema sea sobre el estatus de PR (otro hecho que no debe ser sorpresa). En el tema 4, podemos ver palabras como “Rivera”, “Schatz”, “senado”, “ley”, “electoral” las cuales pueden sugerir que Jay Fonseca hablaba acerca de la reforma electoral. En el tema 5, palabras como “alcalde”, “gobierno”, “pueblo” pudiera sugerir que el Lcdo. Fonseca habla de las alcaldías y municipios. Por último en el tema 6, las palabras “gobernadora”, “Vázquez”, “Wanda”, “salud”, “covid”, “pruebas” es un indicador que Jay tocaba el tema de la respuesta de la exgobernadora al COVID-19.

Mis interpretaciones y/o sugerencias acerca de qué temas toca Jay Fonseca en sus columnas no son 100% perfectas pero si nos sirven para tener una mejor idea cuales palabras utiliza Jay Fonseca en sus columnas y su proporción de utilización conjunto a otras palabras. También quiero enfatizar que este algoritmo no es prueba de que se esta moldeando la opinión pública. Sin embargo quiero expandir en el hecho de que para poder entender cómo se está moldeando la opinión pública hay que tener en cuenta qué palabras se utilizan con mas frecuencia y, a su vez, esas palabras nos ayudan a arrojar luz de los temas que una persona con influencia mediática toca.

Este análisis es una minúscula parte de mi intento de contestar la pregunta de como podemos medir cuantitativamente de si se está moldeando la opinión pública y como. Queda mucha tela por cortar y mucho contexto es requerido para complementar este análisis pero... #LosDatosSonLosDatos

Bibliografia

Topic Modeling. (2021, Mar 16). Retrieved from Text Mining with R: <https://www.tidytextmining.com/topicmodeling.html>