

Reporte Ejecutivo: Conferencias de Prensa de AMLO 2019-2023.



En este reporte analizaremos las conferencias de prensa del presidente Andrés Manuel López Obrador a través de análisis de tópicos y métodos tales como SVD y NMF.

Número Optimo de Tópicos.

A través un análisis de estabilidad usando el criterio de Spearman realizamos una búsqueda probando hasta 35 tópicos, logrando la mejor estabilidad con $k = 15$ tópicos. Posteriormente se realizó una análisis de los resultados con $k = 8, 10, 12, 14, 15, 16$ analizando los modelos resultantes, tanto en SVD como en NMF. Después de consultarlo con un experto en la materia se decidió que los modelos con 12 y 15 tópicos eran mejores.

En base a lo anterior, se opto por un número de 15 tópicos para este análisis.

Resultados SVD.

Con esta metodología se identificaron los tópicos de la tabla 1:

SVD (15)	
Tópico 1	Corrupción
Tópico 2	Salud
Tópico 3	CFE/Energía
Tópico 4	Tren Maya
Tópico 5	COVID/Vacuna
Tópico 6	Seguridad
Tópico 7	EU/RE/Migración
Tópico 8	Desastre Acapulco
Tópico 9	Programas Económicos
Tópico 10	Opositores Políticos
Tópico 11	Educación
Tópico 12	Nuevo Aeropuerto
Tópico 13	Otros
Tópico 14	Economía
Tópico 15	Electoral

Tabla 1: Tópicos SVD

Pese a tener 15 tópicos en la clasificación al momento de clasificarlos algunos no cuentan con ninguna mañanera asignada. Por ejemplo, COVID no se le asigno nada, probablemente pues estos estarían contenidos en el tópico más general, Salud.

El modelo parece tener dificultades en separar el tópico de corrupción, esto podría deberse al amplio uso y generalidad del término. Por otro lado, parece ser bueno separando entre si las demás clases como podemos observar el t-SNE de la figura 1.

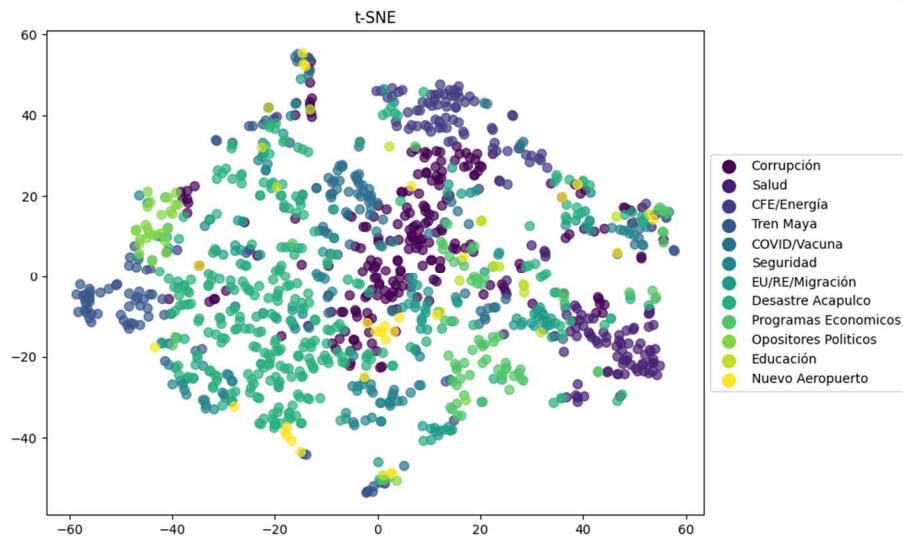


Figura 1. t-SNE con clasificación SVD

Resultados NMF.

Con esta metodología se identificaron los tópicos de la tabla 2:

	NMF (15)
Tópico 1	Corrupción
Tópico 2	Salud
Tópico 3	CFE/Energía
Tópico 4	Tren Maya
Tópico 5	COVID/Vacuna
Tópico 6	Seguridad
Tópico 7	EU/RE/Migración
Tópico 8	Política
Tópico 9	Programas Económicos
Tópico 10	PEMEX/Petróleo
Tópico 11	Educación
Tópico 12	Nuevo Aeropuerto
Tópico 13	Electoral
Tópico 14	Economía
Tópico 15	Desastre Acapulco

Tabla 2: Tópicos NMF

Se puede observar resultado muy similares en cuanto la identificación de los tópicos. En este caso no se presentan problemas de tener tópicos vacíos al momento de clasificar. Los tópicos parecen ser consistentes entre sí y fáciles de interpretar.

Logro una ligera mejor representación en t-SNE

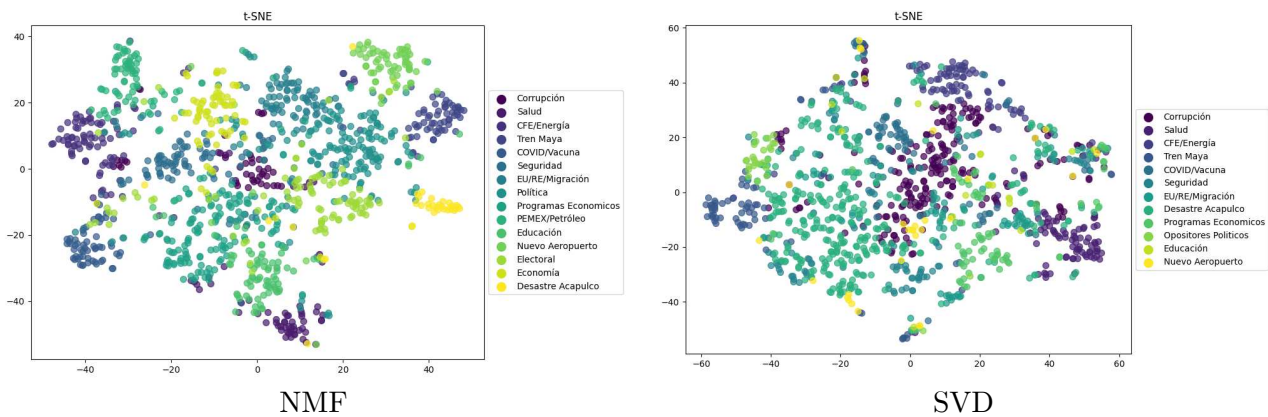


Figura 2. Comparación t-SNE SVD vs NMF

Comparación SVD y NMF.

En general ambos métodos resultan en resultados similares a la hora de detectar tópicos, sin embargo, SVD generaba cierto Tópicos difíciles de clasificar con mayor frecuencia que NMF. Además, NMF no presenta problemas en cuanto a tener tópicos vacíos. Por lo que NMF parece un modelo más fuerte a la hora de detectar y clasificar tópicos. Por otro lado, SVD es un método más general, y suele capturar mejor relaciones entre los datos por su relación con análisis de PCA, puede que sea un método más descriptivo.

Índice para tópicos.

Para desarrollar el índice es necesario tener en cuenta la diversidad como la importancia general pero no global (Pues puede tener mucho sesgo). La manera natural de obtener dicho índice sería de forma frecuentista asignando la frecuencia y luego normalizar o estandarizar. A este enfoque le agregamos la perspectiva general de los tópicos dentro de un mes mediante la frecuencia normalizada durante ese mes obteniendo un modelo inicial:

$$I_{i,j} = \frac{f_{i,j}}{\max_j(f_{i,j})} \times \frac{s_i}{\max_i(s_i)}$$

Donde:

- $f_{i,j}$: Representa la frecuencia de aparición del tópico i en la semana j .
- $\max_j(f_{i,j})$: Es la máxima frecuencia de cualquier tópico durante la semana j , lo que normaliza $f_{i,j}$ en una escala de 0 a 1, donde 1 significa que el tópico fue el más discutido esa semana.
- s_i : Es el puntaje acumulado del tópico i a lo largo de un periodo (mensual), que refleja la importancia general del tópico.
- $\max_i(s_i)$: Es el máximo puntaje acumulado de cualquier tópico en el mismo periodo, utilizado para normalizar s_i .

Esta fórmula combina la importancia particular del tópico en la semana (frecuencia normalizada) y la importancia general del tópico en el período (puntaje normalizado). Un valor alto del índice indica que el tópico es importante tanto en la semana como en el período.

Por otro lado, necesitamos incluir un parámetro de diversidad de los tópicos en cada semana, para obtener la relevancia relativa del tópico. Para esto consideramos el índice de Shannon que mide diversidad de una distribución de probabilidad:

$$H_j = - \sum_{i=1}^n p_i \log(p_i)$$

Donde:

- p_i : Representa la proporción del total de discusiones que el tópico i representa en la semana j .
- n : Es el número total de tópicos diferentes discutidos durante la semana j .
- $\log(p_i)$: Se utiliza el logaritmo natural de p_i , reflejando el aporte de cada tópico a la diversidad total.

Sin embargo, consideramos e^{-H_j} para guardar el impacto del índice positivo y entre 0 y 1. Así el índice de Shannon en el modelo, busca dar más importancia a la localidad semanal.

La propuesta final del índice es la siguiente:

$$I_{i,j} = \frac{f_{i,j}}{\max_j(f_{i,j})} \times \frac{s_i}{\max_i(s_i)} \times e^{-H_j}$$

Observación. Normalizamos mediante mínimo-máximo para lograr una escala comparable entre 0 y 1.

Resultados del análisis.

Los temas principales en las mañaneras del presidente son:

- Corrupción
- Salud
- CFE/Energía
- Tren Maya
- COVID/Vacuna
- Seguridad

Estando presentes prácticamente en todos los modelos independientemente de si se usa SVD o NMF o incluso de los valores de k (para $k \geq 8$ incluso desde antes para alguno tópicos).

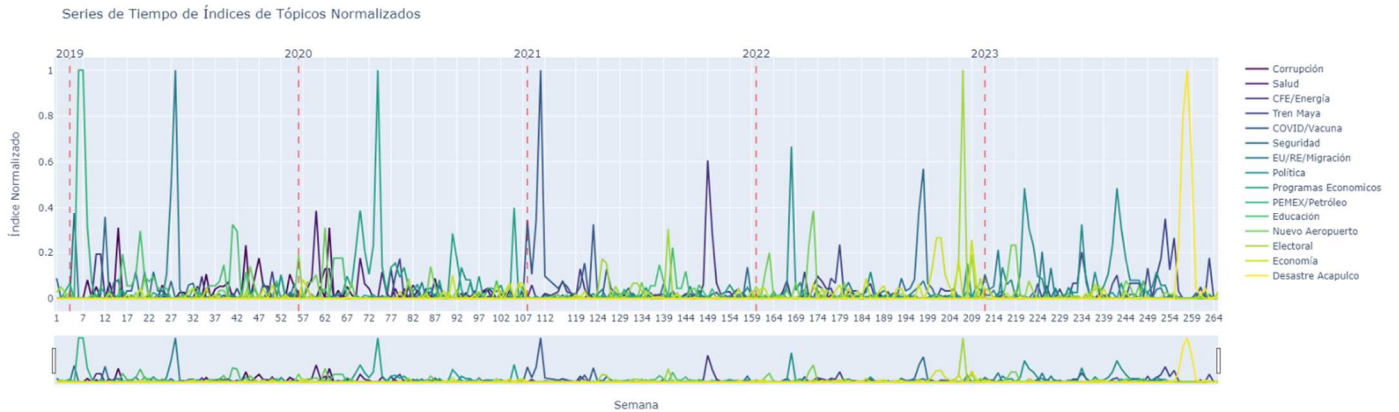


Figura 3. Lineal del tiempo de Tópicos Normalizados 2019-2023

Por otro lado, observando la línea de tiempo resultante del índice preestablecido obtenemos (Recomiendo ver la visualización en [Lineal del tiempo.html](#) para obtener la vista individual de cada tópico):

- Corrupción.

Es evidente que suele hablar más de este tópico en la primera mitad de su sexenio. Podría deberse a que conforme avanza su periodo, los caso de corrupción de sus opositores son menos debido a que su partido es el predominante.

- Salud.

Claramente este tópico predomino durante el año 2020, muy probablemente reflejando la pandemia del CORONAVIRUS.

- CFE.

Tiene su pico a finales del 2022, justo cuando se debate la aprobación de la reforma eléctrica.

- Tren Maya.

Empieza su actividad en la segunda parte de su sexenio y va incrementado hasta su pico en 2023. Sus picos pueden estar relacionados con las protestas ambientalistas relacionadas, así como su inauguración a finales del 2023.

- COVID/VACUNA.

Esta componente refleja las campañas de vacunación contra el COVID-19 llegando a su pico en las primeras semanas de 2021 y bajando paulatinamente.

- Seguridad

El presidente habla más de seguridad en los primeros años de su sexenio, seguramente relacionado con la aprobación de la Guardia Nacional. Hay un pico a finales de 2022.

- EU.

Tiene su pico a mediados de 2019, por las elecciones de los EU.

- Política.

Esta más activo a partir de 2022, relacionado con las reforma electorales, y el inicio de sus secciones contra opositores.

- PEMEX.

Este tópico refleja claramente el problema de abastecimiento y huachicol del inicio de su sexenio.

- Educación.

Un tema recurrente en 2019 y 2021. Relacionado seguramente a sus programas de becas.

- Aeropuerto.

Tópico constante, sus picos iniciales pueden debido a las construcciones y cancelación del anterior. Su pico de 2022 es debido a su inauguración mientras que se menciona recurrentemente en 2023 debido a su polémico rendimiento.

- Desastre Acapulco.

Captura perfectamente lo sucedido por el Huracán Otis.

- Economía.

Un tema recurrente, pero aumenta a finales de 2022, tal vez relacionado con el buen rendimiento de la moneda nacional.

- Entre otras.

Como podemos Observar el modelo parece capturar evento que marcaron puntualmente como generalmente el sexenio del AMLO. Lo que nos permite hacer una radiografía de su gobierno.

Conclusiones.

El gobierno de AMLO ha sido caracterizado por distintos evento entre los que destacan claramente el caso de huachicoleros, el Huracán Otis y el COVID. Sus proyectos claves han sido el Tren Maya y el nuevo Aeropuerto. Le ha dado una importancia muy grande a la CFE y a las becas estudiantiles. Y a partir de su segunda mitad ha disminuido drásticamente su hable de corrupción para cambiarla por un discurso electoral/político contra opositores.

El uso de las técnicas vistas nos permitió realizar este análisis que da un buen panorama del sexenio del presidente. Podríamos mejorar el ejercicio en el caso de SVD agregando rotaciones para maximizar las separaciones como se realizo en mi caso.