

Tarea 1

Universidad Autónoma De Nuevo León
Facultad de Ciencias Físico-Matemáticas
Procesamiento y clasificación de datos

Análisis estadístico de textos

Nombre: Ricardo Luna Escobedo

Matricula: 1805328

Fecha: 09/03/2025

1. Resumen

En este trabajo observaremos el empleo de la programación en python para poder analizar 2 textos de tipo político: el escrito hecho por Lázaro Cárdenas en 1938, con motivo de la expropiación petrolera, y el escrito de Nelson Mandela cuando asumió la presidencia de Sudáfrica en 1994. Si bien, ambos escritos representan acontecimientos históricos diferentes, ambos se caracterizan por representar el nacionalismo.

Las técnicas de análisis de texto se han empleado hasta el momento de manera tradicional (lápiz, papel y lectura), veamos ahora de que manera nos pueden ayudar las nuevas tecnologías en cuanto al aprendizaje de escritos.

2. Introducción

El aprendizaje y la lectura son y siempre serán las armas más poderosas del mundo. Hemos visto que, desde la Grecia clásica, son un tema que constantemente va evolucionando, tanto así que, para asegurarse que la intelectualidad se siguiera preservando a lo largo de los años, todos los conocimientos que se han descubierto han sido plasmados en libros.

La técnica más utilizada para el tema de analizar textos es: el resumen. Con el auge de las nuevas tecnologías, el ser humano ya no solo se limita al resumen de escritos mediante el uso de lápiz y papel, si no que emplea los algoritmos informáticos (python, Rstudio, Java, etc) para agilizar los procesos de lectura y extracción importante de información. Aunque mucha gente siga prefiriendo el método tradicional, lápiz, papel y lectura, nunca está de más probar cosas nuevas y ver que es lo que la tecnología nos puede explicar.

3. Investigación

3.1. Datos

La historia universal nos ha dejado grandes líderes, tanto políticos como humanitarios, como: Benito Juárez, Mahatma Gandhi, Simón Bolívar, etc. Pero de entre todos ellos, mi interés personal surgió para escribir acerca de Lázaro Cárdenas y Nelson Mandela.

Lázaro Cárdenas del Río fungió como presidente de México entre 1934 y 1940. Su acción más memorable: la expropiación del petróleo mexicano en 1938. Por otra parte, Nelson Rolihlahla Mandela fungió como presidente de Sudáfrica entre 1994 y 1999. Su acción más memorable: la conciliación racial nacional.

Ambos personajes obraron y vivieron de manera muy diferente, a pesar de que ambos eran políticos. Pero había algo que compartían en común, además de ser presidentes, y era: su enorme nacionalismo. El nacionalismo puede ser interpretado de manera subjetiva, ya que muchas personas luchan por su país teniendo objetivos diferentes a través de ideologías particulares. En esta investigación analizamos como era la ideología de cada presidente a partir de los discursos que dieron una vez que completaron su acción más memorable. El análisis se hizo con el uso del lenguaje de programación python.

3.2. Descripción

1. Discursos (parciales)

1.1. Lázaro Cárdenas

La actitud asumida por las compañías petroleras negándose a obedecer el mandato de la Justicia Nacional que por conducto de la Suprema Corte las condenó en todas sus partes a pagar a sus obreros el monto de la demanda económica que las propias empresas llevaron a los tribunales judiciales por inconformidad con las resoluciones de los Tribunales del Trabajo, impone al Ejecutivo de la Unión el deber de buscar en los recursos de nuestra legislación un medio eficaz que evite definitivamente, para el presente y para el futuro, el que los fallos de la justicia se nulifiquen o pretendan nulificarse por la sola voluntad de las partes o de alguna de ellas mediante una simple declaratoria de insolvencia como se pretende hacerlo en el presente caso, no haciendo más que incidir con ello en la tesis misma de la cuestión que ha sido fallada... [1]

1.2. Nelson Mandela

En el día de hoy, todos nosotros, mediante nuestra presencia aquí y mediante celebraciones en otras partes de nuestro país y del mundo, conferimos esplendor y esperanza a la libertad recién nacida. De

la experiencia de una desmesurada catástrofe humana que ha durado demasiado tiempo debe nacer una sociedad de la que toda la Humanidad se sienta orgullosa.

Nuestros actos cotidianos como sudafricanos normales deben producir una realidad sudafricana actual que reafirme la creencia de la Humanidad en la justicia, refuerce su confianza en la nobleza del alma humana y dé aliento a todas nuestras esperanzas de una vida gloriosa para todos. Todo esto nos lo debemos a nosotros mismos y se lo debemos a los pueblos del mundo que tan bien representados están hoy aquí... [2]

2. Palabras promedio

El promedio de palabras por párrafo para el discurso de Nelson Mandela es: 29.89

El promedio de palabras por párrafo para el discurso de Lázaro Cárdenas es: 73.90

Lázaro Cárdenas incluyó, en promedio, más palabras por párrafo en su discurso que Nelson Mandela.

3. Tokenización

El análisis de texto involucra el estudio de palabras relevantes; aquellas palabras que consideramos no relevantes se les denomina: *Stopwords*.

Ejemplo en una oración: La actitud asumida por las compañías petroleras. Al eliminar "Stopwords" solo nos quedaríamos con las palabras: actitud, asumida, compañías y petroleras. Este proceso se aplicó para todos los textos, de tal forma que consideremos aquellas palabras que nos pueden ofrecer contexto significativo. Una vez hecha, convertimos en listas las palabras restantes para poder manipularlas de manera más eficiente e interactiva.

4. Gramática

Para poder agregar más contexto a las palabras (tokenizadas), de nuestro análisis, se tienen 2 técnicas a emplear: stemming y lematización. El stemming busca la raíz de las palabras a través de los morfemas, mientras que la lematización busca la raíz mediante sinónimos o conjugaciones. Se utilizó la lematización porque observamos que, a diferencia del stemming, toma palabras completas y no solamente prefijos y sufijos. Vease la Figura 1.

Figura 1: Lematización de discursos

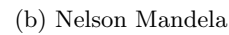
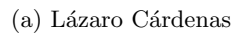
	Palabra Original	Lematización (spaCy)		Palabra Original	Lematización (spaCy)
0	actitud	actitud	0	día	día
1	asumida	asumido	1	hoy	hoy
2	compañías	compañía	2	mediante	mediante
3	petroleras	petrolero	3	presencia	presencia
4	negándose	negar él	4	aquí	aquí
...
1853	ignacio	ignacio	500	miedo	miedo
1854	garcía	garcía	501	presencia	presenciar
1855	téllez	téllez	502	automáticamente	automáticamente
1856	conocimiento	conocimiento	503	libera	liberar
1857	directo	directo	504	demás	demá

(a) Lázaro Cárdenas

(b) Nelson Mandela

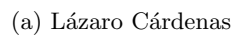
Ahora que hemos adaptado nuestra información a un nivel procesable para manejar en python, podemos proceder a realizar análisis más detallados.

Figura 2: Gráficas de conteo palabras



La descripción anterior de las palabras también las podemos ver en los recursos WordCloud...

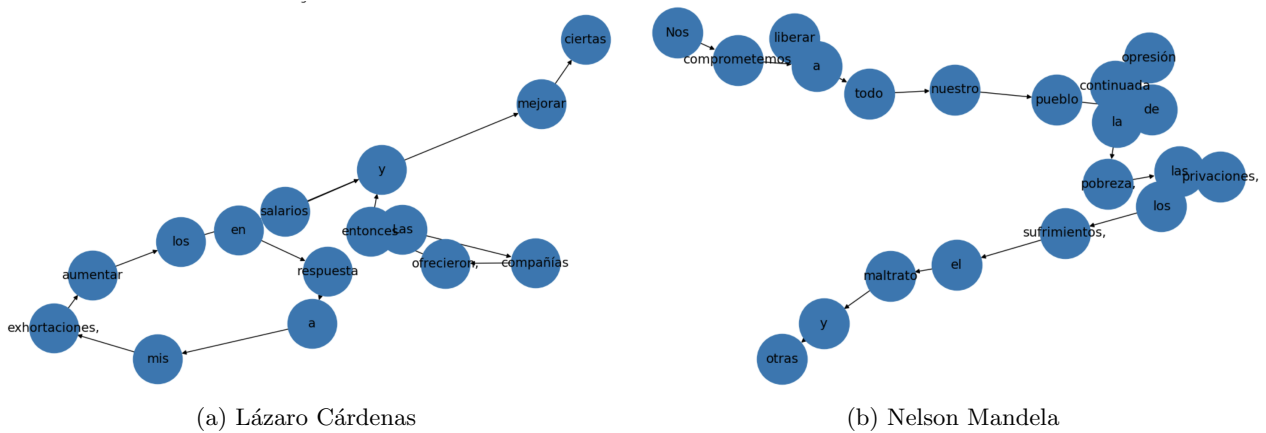
Figura 3: Gráficas de representación léxica



Del mismo modo, podemos ver la misma relación mediante un gráfico n-gramas...

5.3. N-gramas

Figura 4: Gráficas de trigramas



Para el discurso de Lázaro Cárdenas, el trigramas nos añade información sobre como se exhorta al aumento de salarios hacia los trabajadores por parte de las compañías petroleras. Mientras que con Mandela se exhorta a no volver a permitir que el pueblo negro Sudafricano sufra opresión, pobreza o maltrato.

4. Conclusión

Siento una gran admiración por personas de índole nacionalista. Es por esa razón que decidí realizar este artículo acerca de 2 personas que han formado parte no solo de la historia política, sino de la historia humanitaria. Lázaro Cárdenas, en su momento, luchó por proteger los intereses de los recursos mexicanos contra los extranjeros; en este caso: el petróleo. Nelson Mandela, por su parte, luchó por proteger y defender los derechos de las personas negras Sudafricanas contra las injusticias que la gente blanca les impuso durante la época del Apartheid. Con el análisis realizado, podemos darnos cuenta de la gran utilidad que fue la tecnología/programación con Python para estudiar ambos discursos; pudimos obtener una idea de que tratan los escritos sin necesidad de leerlos completamente. También soy una persona que tiene un gran gusto en temas de lectura. Fue muy interesante ver la comprensión lectora desde un punto de vista más algorítmico que del tradicional.

5. Referencias

- [1] Lázaro Cárdenas del Río. Discurso con motivo de la expropiación petrolera, 1938. Mensaje a la nación, 18 marzo 1938.
- [2] Nelson Rolihlahla Mandela. Discurso de toma de posesión, 1994. Mensaje a la nación, 10 mayo 1994.