

Tarea 2

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

Análisis de sentimientos

Nombre: Ricardo Luna Escobedo

Matricula: 1805328

Fecha: 03/04/2025

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. También notaremos una nueva técnica utilizada para poder conocer más a detalle el contenido; no solo de manera numérica, sino de manera emocional.

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.

Dentro de los algoritmos informáticos tenemos el desarrollo de una nueva área tecnológica: inteligencia artificial. Este tema ha estado mucho últimamente en boca de todos porque es una ciencia que busca desarrollar el comportamiento humano a partir de la computación. En este ensayo haremos uso de una librería en python hecha a base de la inteligencia artificial: **análisis de sentimientos**. Ésta técnica de procesamiento de lenguaje natural, nos permite identificar y extraer opiniones o emociones expresadas en un texto. El objetivo principal es clasificar sentimientos generales del texto como positivos, negativos o neutrales.

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)
0	actitud	actitud
1	asumida	asumido
2	compañías	compañía
3	petroleras	petrolero
4	negándose	negar él
...
1853	ignacio	ignacio
1854	garcía	garcía
1855	téllez	téllez
1856	conocimiento	conocimiento
1857	directo	directo

(a) Lázaro Cárdenas

	Palabra Original	Lematización (spaCy)
0	día	día
1	hoy	hoy
2	mediante	mediante
3	presencia	presencia
4	aquí	aquí
...
500	miedo	miedo
501	presencia	presenciar
502	automáticamente	automáticamente
503	libera	liberar
504	demás	demá

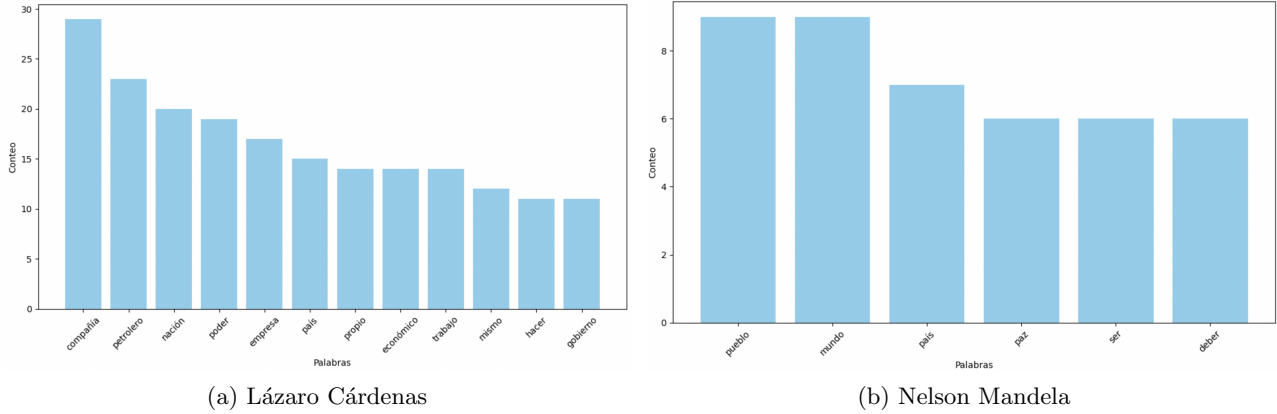
(b) Nelson Mandela

5. Estadísticas descriptivas

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

5.1. Gráficas barras

Figura 2: Gráficas de conteo palabras



Aquí podemos ver como en el discurso de Lázaro Cárdenas se detectaron más palabras relevantes debido a que éste es más largo que el de Mandela. Entre las palabras que más se repitieron, para el caso de Cárdenas, se encuentran aquellas relacionadas con las compañías, el petróleo, la nación, el poder, el trabajo, el gobierno, la economía, etc. Mientras que para el caso de Mandela, se menciona más a la paz y el deber. Aunque hay mucha diferencia entre la temática de ambos discursos, podemos ver que hay un elemento en común: el nacionalismo.

6. Análisis de sentimientos

Para propósitos más prácticos, y entendibles, los discursos se dividieron en fragmentos. De tal manera que se va asignar una probabilidad positiva, negativa o neutral a la oración analizada para determinar la expresión escrita; el sentir del orador/escritor.

Comencemos con el discurso de Nelson Mandela.

Figura 3: El expresar Mandela

Fragmento	Texto	Sentimiento	Prob_POS	Prob_NEU	Prob_NEG
1	¡ÚEn el d/a de hoy, todos nosotros, mediant	NEU	0.3197420835494995	0.3903273642063141	0.2899306118488312
2	Nuestros actos cotidianos como sudafricanos	POS	0.799877942	0.17709000408649445	0.023032056
3	Sin la menor vacilaci n digo a mis compatrio	POS	0.9156439304351807	0.078276202	0.006079793
4	Nosotros, el pueblo sudafricano, nos sentimos	POS	0.907432497	0.073941641	0.018625928089022636
5	Apreciamos hondamente el papel que para lle	NEU	0.2536860704421997	0.4088522493839264	0.3374616801738739
6	Ha llegado el momento de curar las heridas.	POS	0.5431661009788513	0.42187315225601196	0.034960736
7	Ha llegado el momento de salvar los abismos c	NEU	0.37200722098350525	0.4381718337535858	0.18982090055942535
8	Al fin hemos logrado la emancipaci n polí t	POS	0.7500255107879639	0.19562581181526184	0.054348621517419815
9	Como muestra de este compromiso de renova	POS	0.6385073065757751	0.2800644040107727	0.081428327
10	Sus sue os se han cumplido. La libertad es si	POS	0.8967949151992798	0.097693525	0.005511589
11	Nos sentimos humildes y enaltecidos por el ho	POS	0.8862369656562805	0.080641404	0.033121616
12	Somos conscientes de que el camino hacia la	POS	0.5025399923324585	0.4382060766220093	0.059253942
13	Sabemos muy bien que ninguno de nosotros pc	NEG	0.055279616	0.4006231129169464	0.5440973043441772
14	Por eso debemos actuar juntos como un puebl	POS	0.7212249636650085	0.22713227570056915	0.051642779260873795
15	Que haya justicia para todos. Que haya paz pa	POS	0.6575050354003906	0.3034524917602539	0.039042502641677856
16	Que haya trabajo, pan, agua y sal para todos.	POS	0.6198028922080994	0.3264319598674774	0.053765177726745605
17	Que seamos conscientes de que nuestros cue	POS	0.8167296051979065	0.17213386297225952	0.011136498302221298
18	Nunca, nunca jam s debe ocurrir que esta he	NEG	0.007369241	0.028473680838942528	0.9641570448875427
19	Que reine la libertad.	POS	0.5888569951057434	0.3650021255016327	0.046140838
20	El sol jam s se pondrá sobre un logro hum	POS	0.8315756320953369	0.13080739974975586	0.037617017
21	Muchas gracias,Áu.	POS	0.9026739597320557	0.081696257	0.015629788860678673
22	Nuestro miedo m s profundo no es el de ser i	NEG	0.017459966242313385	0.14082355797290802	0.841716468
23	Es nuestra luz, no nuestra oscuridad, lo que no	NEG	0.039036967	0.18274350464344025	0.7782195806503296
24	Nos preguntamos ¿qui soy yo para ser br	NEU	0.1291360855102539	0.48727044463157654	0.38359346985816956
25	Eres hijo del universo.	POS	0.7574295997619629	0.2050824910402298	0.037487913
26	El hecho de jugar a ser peque o no le sirve al	NEG	0.00732213	0.063759387	0.9289184212684631
27	No hay nada iluminador en encogerse para que	POS	0.614014745	0.3069031834602356	0.079082102
28	Nacemos para poner de manifiesto la gloria de	POS	0.8848521113395691	0.10464798659086227	0.010499863885343075
29	No est solamente en algunos de nosotros, e	NEU	0.2924775779247284	0.6300896406173706	0.077432826
30	Y mientras dejamos lucir nuestra propia luz, in	POS	0.5835754871368408	0.3937273919582367	0.022697173058986664

El discurso fue dividido en 30 fragmentos con las siguientes características:

- 20 fragmentos categorizados como: positivos.
- 5 fragmentos categorizados como: neutros.
- 5 fragmentos categorizados como: negativos.

Podemos ver que la categoría que recibió más asignaciones fue la positiva; tuvo la mayor cantidad de probabilidades altas. Por lo que podemos concluir que el discurso de Mandela dirige un mensaje inspiratorio; ya que lo que busca es brindar esperanza a su público. Pero observemos también el contenido de los fragmentos:

- Los fragmentos categorizados como positivos recibieron esa asignación porque incluyen palabras como: reafirmar, creencia, justicia, confianza, nobleza, hermoso país, renovación personal, júbilo, regocijo, etc.
- Los fragmentos categorizados como neutros recibieron esa asignación porque incluyen palabras como: sociedad orgullosa, transición, protección, dividir, construir, solamente, algunos de nosotros, etc.
- Los fragmentos categorizados como negativos recibieron esa asignación porque incluyen palabras como: miedo, inapropiado, ninguno, soledad, oscuridad, asustar, nunca jamás, opresión, indignidad, escoria, etc.

Ahora sigamos con el discurso de Lázaro Cárdenas.

Figura 4: El expresar Cárdenas

Fragmento	Texto	Sentimiento	Prob_POS	Prob_NEU	Prob_NEG
1	La actitud así NEG		0.004445201	0.069264501	0.9262903928756714
2	Hay que con NEG		0.003717122	0.029055033	0.9672278165817261
3	Por otra parte NEG		0.004759998	0.034906834	0.9603331685066223
4	Ya en estas c NEG		0.005121328	0.071774907	0.9231036901473999
5	Y en esta situ NEG		0.003477709	0.031622525	0.9649004340171814
6	Es evidente q NEG		0.005812225	0.047096624970436096	0.9470911622047424
7	Se trata de ur NEG		0.002899502	0.023394506	0.9737059473991394
8	En tal virtud s NEU		0.2578297257423401	0.520921528	0.22124874591827393
9	La historia de NEU		0.22518521547317505	0.6932109594345093	0.081603907
10	El a v o de l NEU		0.048125166445970535	0.7997056245803833	0.15216918289661407
11	En junio de l NEU		0.041402407	0.8091417551040649	0.14945587515830994
12	A finales de l NEU		0.086258024	0.8433482646492139	0.070393719
13	Con objeto d NEU		0.097948767	0.8265435099601746	0.075507782
14	El Sindicato d NEU		0.032151505	0.6744954586029053	0.2933530211448694
15	En atenci n NEU		0.038107067346572876	0.7879762649536133	0.17391657829284668
16	Las continger NEG		0.003103491	0.041389763	0.9555068016052246
17	Las compa s NEU		0.029904285445809364	0.6222472786903381	0.34784844517707825
18	En virtud de l NEU		0.32373350858688354	0.6251351833343506	0.0511131327
19	La comisi n NEU		0.037297592	0.6354283094406128	0.3272741734981537
20	Sin embargo r NEG		0.026607414707541466	0.4704533815383911	0.5029391646385193
21	En todas y ca NEU		0.12834899127483368	0.7462329864501953	0.1254180371761322
22	Es por tanto r NEU		0.11723046749830246	0.6356808543205261	0.24708865582942963
23	Para mayor j NEU		0.4802350103855133	0.4511815905570984	0.068583384
24	Se ha dicho h NEU		0.35799068212509155	0.4507322311401367	0.19127705693244934

(a) Parte 1

25	Esta afirmaci NEG		0.009495613	0.072635889	0.917868435
26	Riqueza pote NEG		0.07214871	0.2878909707069397	0.6399603486061096
27	Examinemos NEG		0.009107332	0.2368663102388382	0.7540263533592224
28	En cu l c NEG		0.003567889	0.046012416	0.9504196047782898
29	Qu n no NEG		0.002514992	0.016933688893914223	0.9805512428283691
30	Otra conting NEG		0.004212496	0.035054199397563934	0.9607334136962891
31	Nadie discute NEG		0.008067944	0.057754226	0.9341778755187988
32	Es por lo tanti NEG		0.011313599534332752	0.086882904	0.9018035531044006
33	Planteadas as NEG		0.14554433524608612	0.36025646328926086	0.4941991865634918
34	El Gobierno h NEU		0.2586427927017212	0.5898951888084412	0.15146195888519287
35	Sin embargo, NEU		0.1736033856868744	0.5021700263023376	0.32422658801078796
36	Es preciso qu POS		0.631702721	0.31782782077789307	0.050469458
37	Y como pudie NEG		0.011824940331280231	0.078552574	0.9096224308013916
38	En tal virtud y NEG		0.007024982	0.099401757	0.8935732245445251
39	Considerand NEG		0.008918999	0.099982709	0.8910982012748718
40	Por lo expues NEU		0.11745361983776093	0.7907688909799194	0.091777466
41	DECRETO. NEU		0.2298949509859085	0.5801103711128235	0.189994678
42	Art culo 1 NEU		0.15382032096385956	0.683684766	0.16249488294124603
43	Art culo 2 NEU		0.080933407	0.8233891725540161	0.095677361
44	Art culo 3 NEU		0.075839102	0.8327733874320984	0.091387518
45	Art culo 4 NEU		0.04523531	0.8429820673370361	0.11177262663841248
46	Este decreto NEU		0.11775260418653488	0.8132873177528381	0.068960041
47	Presidente NEU		0.16150332987308502	0.7018563747406006	0.13664023578166962

(b) Parte 2

El discurso fue dividido en 47 fragmentos con las siguientes características:

- 2 fragmentos categorizados como: positivos.
- 24 fragmentos categorizados como: neutros.
- 21 fragmentos categorizados como: negativos.

Podemos ver que la categoría que recibió más asignaciones fue la neutra; tuvo la mayor cantidad de probabilidades altas. Por lo que podemos concluir que el discurso de Cárdenas dirige un mensaje de concientización con un toque de enojo; ya que lo que busca es hacerle ver al público la importancia de proteger los recursos de la nación contra intereses extranjeros. Pero observemos también el contenido de los fragmentos:

- Los fragmentos categorizados como positivos recibieron esa asignación porque incluyen palabras como: franco, optimismo, espíritu, salvar, esfuerzo, etc.
- Los fragmentos categorizados como neutros recibieron esa asignación porque incluyen palabras como: procurar, convenio, aquiescencia, diverso, conciliatorio, amparo, etc.
- Los fragmentos categorizados como negativos recibieron esa asignación porque incluyen palabras como: negarse, condena, inconformidad, dificultades, descontento, etc.

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, y de entender su expresar de manera probabilística a través del uso del análisis de sentimientos. 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.