

Minería de datos aplicada a los discursos presidenciales de Pedro Castillo Terrones en Perú

Data mining applied to the presidential speeches of Pedro Castillo Terrones in Perú

Chung Pinzás, A. R. y Inche Mitma, J. L.



Alfonso Ramón Chung Pinzás, Universidad Nacional Mayor de San Marcos (Perú)

Doctor en Ingeniería Industrial. Es profesor principal de la Universidad Nacional Mayor de San Marcos. Sus principales líneas de investigación están relacionadas a la minería de datos, big data, ciencia de datos e inteligencia artificial.

<https://orcid.org/0000-0002-8073-2396>, achungp@unmsm.edu.pe



Jorge Luis Inche Mitma, Universidad Nacional Mayor de San Marcos (Perú)

Doctor en Ciencias Administrativas. Es profesor principal de la Universidad Nacional Mayor de San Marcos. Sus principales líneas de investigación están relacionadas a la minería de datos, emprendimiento y gestión del conocimiento.

<https://orcid.org/0000-0002-5887-8591>, jinchem@unmsm.edu.pe

Recibido: 26-09-2023 – Aceptado: 25-01-2024

<https://doi.org/10.26441/RC23.1-2024-3417>

RESUMEN: El artículo presenta los resultados de una investigación no experimental, descriptiva y explicativa. Dicha investigación está relacionada a la minería de texto y de sentimientos, para ello se trabajó con la totalidad de los discursos oficiales del expresidente del Perú, Pedro Castillo Terrones, los que ascienden a 12 y son de dominio público en el portal del estado peruano. El objetivo fue determinar cuál es el mensaje detrás de los discursos antes mencionados. Los datos fueron procesados en R a través de RStudio, y el lexicón NRC. Los resultados arrojan mensajes orientados a transmitir la idea de un estado benefactor; por lo tanto, no están orientados al fomento de la iniciativa emprendedora en la población, sino más bien a la recepción por parte del estado de los recursos, en especial el económico, esto acorde con el ideario del partido que lo llevó al poder. En cuanto a los sentimientos, predominan los positivos. Al ser un tema con una alta carga política e ideológica, se ha buscado aislar dichas cargas, restringiendo la interpretación a los resultados antes mencionados.

Palabras clave: análisis del discurso; minería de sentimientos; riqueza; minería de texto; estado.

ABSTRACT: The article presents the results of a non-experimental, descriptive and explanatory research. This research is related to the mining of text and feelings, for this we worked with all the official speeches of the former president of Peru Pedro Castillo Terrones, which amount to 12 and are in the public domain on the Peruvian state portal. The objective was to determine what the message is behind the before mentioned speeches of the former president. The data were processed in R through RStudio, and the lexicon NRC. The results show messages aimed at transmitting the idea of a welfare state; Therefore, they are not aimed at promoting entrepreneurship, initiative in the population, but rather at the state's receipt of resources, especially economic resources, which is in line with the ideology of the party that brought it to power. As for feelings, positive ones predominate. Being a topic with a high political and ideological charge, an attempt has been made to isolate these charges, restricting the interpretation to the before mentioned results.

Keywords: speech analysis; sentiment mining; wealth; text mining; state.

1. Introducción

El presente artículo aborda una investigación del tipo descriptiva y explicativa, se enfoca en analizar el contenido de los discursos del expresidente Pedro Castillo Terrones, buscando determinar cuál es el mensaje que ha intentado transmitir a los peruanos en el conjunto de todas sus intervenciones oficiales, pero desde un punto de vista estadístico y por lo tanto cuantitativo en el marco de la minería de datos, de tal forma que se pueda abordar el tema desde una perspectiva distinta al análisis político tradicional.

El problema de investigación a tratar es el mensaje detrás de los discursos oficiales del expresidente Pedro Castillo Terrones, y el objetivo es determinar cuál es el mensaje detrás de los discursos oficiales del expresidente Pedro Castillo Terrones.

La investigación presentada en el este artículo se justifica toda vez que es necesario conocer el verdadero mensaje que trata de comunicar un político, más aún si se trata de un presidente, de tal forma que las personas puedan ejercer mejor su condición de ciudadanos, discrepando o apoyando si fuera el caso, asimismo, también para poder analizar si es que lo que dice se condice con lo que hace, es decir, si el mensaje detrás de los discursos está a la par con las acciones de gobierno o en realidad es un mecanismo distractor que vende un escenario que no existe.

Asimismo, la aplicación de la minería de datos a los discursos presidenciales es un tema novedoso que constituye un importante aporte tanto para la reflexión ciudadana como para el análisis político, debido a que en ambos casos proporciona una herramienta estadística de análisis altamente robusta que permite llevar a cuantitativo, diversos aspectos que pueden considerarse como cualitativo.

Por otro lado, el abordaje de este tema tiene una alta carga política e ideológica, sin embargo, una de las ventajas del uso de herramientas como la minería de datos es que permite traducir a números y relaciones estadísticas aspectos que pueden ser interpretados según el analista, por lo tanto, se ha procurado en la medida de lo posible aislar las cargas ideológicas, tomando en cuenta la interpretación de resultados en función a los valores obtenidos.

2. Marco referencial

2.1. Minería de datos

La minería de datos se puede considerar como un elaborado proceso en donde un analista tiene como objetivo identificar en un conjunto de datos, patrones, correlaciones, anomalías etc. (Emigdio, 2022), por otro lado, Beynon (2018) puntualiza que esta herramienta se utiliza con almacenes o mercado de datos, de tal forma que se puedan obtener beneficios de los mismos, mediante la extracción de patrones y correlaciones entre dichos datos, a través del uso de algoritmos automáticos. Para lograr su cometido, la minería de datos se basa en diversas ciencias tales como las matemáticas, estadística, inteligencia artificial (Ortega, 2022) etc.

De hecho, Quiroga, et al., (2022) mencionan que en líneas generales la minería de datos pasa por las siguientes etapas: selección, preprocesamiento, transformación, aplicación de la minería de datos, interpretación y evaluación; de tal forma que al final se llegue a un determinado conocimiento. También es importante destacar que esta herramienta se divide principalmente en minería de texto, de sentimiento y de web, todo esto considerando la fuente de datos a analizar (Ortega, 2022). Para efectos del artículo, se va a considerar tanto la minería de texto como de sentimiento.

2.2. Minería de Texto

La minería de texto constituye una rama de la minería de datos que se relaciona con el texto y busca descubrir información que no figura literalmente en los documentos, capturando temas,

conocimientos claves y relaciones ocultas (Quiroga y otros, 2022), de forma similar, Guerra (2023) menciona que esta herramienta se orienta al descubrimiento de patrones y tendencias mediante el uso de técnicas como por ejemplo el aprendizaje estadístico de patrones; así pues la minería de texto tiene como objetivo transformar, mediante algoritmos, la información textual en números (Torres, 2020).

Complementando lo anterior, se puede definir también como una colección de técnicas que permiten extraer conocimiento a partir de un corpus textual conformado por datos no estructurados, por lo tanto, se inscribe en el campo de la lingüística computacional y del procesamiento de textos (García, 2017), como se puede observar, la minería de texto intenta descifrar lo que hay detrás de las palabras, es decir, cual es el discurso oculto detrás del texto.

Según González & Ramírez (2018) las áreas de la minería de texto son: clasificación automatizada de texto considerando categorías determinadas, clúster de texto, resumen automatizado, extracción de temas y tendencias de los textos. En cuanto a sus fases, se tiene principalmente tres actividades (Universidad de Malaga, s.f.):

- a. **Preprocesamiento de datos:** en esta primera fase, los textos se transforman en un tipo de representación ya sea estructurada o semiestructurada de tal forma que permita su posterior análisis con el consiguiente reconocimiento de los tokens o unidades gramaticales esenciales, con lo cual se organiza el texto como un vector de palabras.
- b. **Descubrimiento:** en esta segunda fase se realiza el análisis de las representaciones anteriormente mencionadas a fin de descubrir patrones u otro conocimiento clave.
- c. **Visualización:** en esta fase final, se presentan los resultados de tal forma que puedan ser explorados por los usuarios.

Sobre las herramientas informáticas disponibles, Joyanes (2020) menciona que son muy populares y que pueden ser parte de *suites* o como software individual; como ejemplos se tiene las librerías especializadas de R y Python.

2.3. Minería de sentimientos

La minería de sentimientos es básicamente una técnica para el procesamiento del lenguaje natural orientado a la identificación del componente emocional detrás de un cuerpo de texto, utilizando para ello la minería de datos y *Machine Learning*, con el fin de buscar sentimientos e información subjetiva (Orozco y otros, 2021).

Asimismo, para Fernández (2019) la minería de sentimiento busca categorizar el valor del sentimiento en dos categorías mutuamente excluyentes tales como “positivo” o “negativo”, aunque en algunos casos puede incluir el término “neutro”; sin embargo, tal y como lo menciona Boden (2017), toda vez que se trata de una valoración es importante analizar el contexto en que se mencionan las palabras.

Dentro de los enfoques para el este tipo de análisis, Alaminos (2023) identifica los siguientes:

- a. **El enfoque basado en lexicón:** en este caso, se toma como base un lexicón predefinido de palabras asociadas a emociones o sentimientos negativos o positivos, este lexicón se utilizará para comparar, clasificar y contar la cantidad de palabras del texto según el tipo de sentimiento al cual pertenecen en el mencionado lexicón; ejemplos de lexicones de sentimientos son los siguientes AFINN, NRC, SentiWordNet y VADER.
- b. **El enfoque basado en aprendizaje automático:** como su propio nombre lo indica, en este caso se utilizan algoritmos de aprendizaje no supervisado, con conjuntos de datos de entrenamiento y prueba, en cuanto a los algoritmos utilizados, se puede mencionar la regresión logística, las máquinas de vectores de soporte, árboles de decisión y redes neuronales.

- c. **El enfoque híbrido:** en este enfoque se combinan ambos enfoques anteriormente mencionados a fin de lograr una mayor robustez del modelo.

En cuanto a sus fases, estas son en parte similares a las mencionadas en el acápite de minería de texto, así pues, se tiene lo siguiente (Rojas, 2022):

- a. Exploración del *dataset*.
- b. Pre-procesamiento.
- c. Exploración y visualización.
- d. Análisis de sentimientos.

2.4. La minería de datos y su aplicación en la comunicación política

La minería de datos al no provenir de las ciencias sociales, su aplicación en la política deviene en un proceso de aprendizaje (Merlo, 2022), ahora bien, cuando se habla sobre este tema, se piensa automáticamente en la campaña presidencial del 2016 que llevó al poder a Trump, sin embargo, el verdadero precursor fue Obama, quien, de la mano con Google, logró la victoria en el 2008 (Muñoz, 2023).

Pero ¿Qué tanto se puede aplicar la minería en el aspecto político?, de hecho, tal y como lo menciona Cotino (2022), versiones sofisticadas de estas herramientas pueden vincular a ciudadanos con diversas ideologías, modelar la conducta de los votantes, etc., por lo tanto, no es de extrañar el inmenso caudal de posibilidades que presenta la minería de texto en el aspecto político.

De hecho, estas posibilidades en política crecen enormemente ya que, asociadas a la inteligencia artificial (IA), es posible escrutar las emociones y sentimientos de la población, más allá del estudio de los gestos y la comunicación no verbal, es por eso por lo que gran cantidad de investigadores exploran modelos de IA para estos fines (Gutiérrez, 2023).

Para la investigación presentada en el este artículo, cabe mencionar que la minería de datos se basa en el principio de que todos dejan una huella digital al visitar las páginas webs, las mismas que pueden ser galletas, IP etc. Posteriormente, estas huellas son almacenadas por los servidores y luego analizadas mediante dicha herramienta (Dorantes, 2023), para el análisis político, su aplicación se fundamenta en que estas huellas se encuentran dentro de los discursos (también se puede ampliar a redes sociales), esto se traduce en frecuencia de palabras, sentimiento de estas, etc., las cuales pueden ser previamente limpiadas y transformadas en vectores para su posterior análisis.

2.5. El contexto de la investigación

Pedro Castillo Terrones fue presidente del Perú por el partido Perú Libre, de orientación Marxista Leninista Mariateguista (Cerrón, 2020), durante el periodo que comprende del 28 de julio del 2022 al 7 de noviembre del 2023, como se puede apreciar, no cumplió su periodo de cinco años de gobierno debido a que fue vacado por el congreso por “incapacidad moral” al amparo del artículo 115 de la constitución política del Perú (Centro de Noticias del Congreso, 2022), produciéndose con ello la inmediata sucesión presidencial por parte de Dina Ercilia Boluarte Zegarra.

Su corto periodo de gobierno, que es en donde se desarrolla la investigación del presente artículo, comienza a gestarse al finalizar la primera vuelta de las elecciones en el Perú, cuando obtiene el 18.92% sobre el 13.40% de su rival Keiko Fujimori del partido Fuerza Popular (El Comercio, 2021), seguidos por otros candidatos. Posteriormente lograría el triunfo en la segunda vuelta con un estrecho margen de 50.126% contra 49.874% de su rival de Fuerza Popular (Jurado Nacional de Elecciones, 2021).

Las elecciones presidenciales del 2021 en el Perú, se caracterizaron primeramente porque se desarrollaron en época de pandemia, con todos los estragos que dicha etapa funesta significó tanto para el Perú como para el mundo, por otro lado se tiene la atomización de la oferta electoral de aquel entonces, de hecho se presentaron 18 candidatos, asimismo, habría que sumarle el exacerbado antifujimorismo, el cuál perjudicaría directamente a la candidata Keiko Fujimori del partido Fuerza Popular (hija del expresidente Alberto Fujimori).

Como se mencionó anteriormente, Pedro Castillo con 18.92% y Keiko Fujimori 13.40% pasan a segunda vuelta, pero como se puede observar los porcentajes de votación, a pesar de ser los dos más altos de entre todos los candidatos, en realidad son a la vista bajos, es decir, la votación lograda en ninguno de los casos era contundente.

Esta segunda vuelta se caracterizó por ser una de las más polarizadas, esta polarización se centró en Pedro Castillo como el representante del zona rural e indígena y Keiko Fujimori, como la representante de la política tradicional, limeña y centralista (Miyagusuku, 2022), asimismo, diversos personajes ligados a la izquierda peruana apoyaron la candidatura de Pedro Castillo, como ejemplo se puede mencionar a Verónica Mendoza (candidata a la presidencia por Juntos por el Perú) quien en una entrevista a una emisora radial mencionó lo siguiente “Si tengo dudas respecto al señor (Pedro) Castillo, de la señora Fujimori tengo certezas. Y tengo certezas de que, con ella, ni a la esquina” (La República, 2021), sin embargo, cabe mencionar dicha candidata, anteriormente también había apoyado a un candidato, en aquella ocasión de la derecha peruana, en el 2016 (Pedro Pablo Kuczynski y su partido PPK) a fin de evitar la presidencia de Keiko Fujimori (TVPerú, 2016).

Otro personaje vinculado a la izquierda peruana que manifestó su rechazo hacía Keiko Fujimori y por ende su apoyo a la candidatura de Pedro Castillo fue el fiscal José Domingo Pérez, quien en una declaración a un medio de comunicación manifestó lo siguiente: “Es un caso inédito que la jefa de la organización criminal que ha acusado el Ministerio Público pueda ocupar la primera función pública” (La República, 2021).

Por el lado contrario, es decir los que estaban a favor de Keiko Fujimori, se le acusó a Pedro Castillo de ser comunista y de pretender instaurar una dictadura del tipo Venezuela (Miyagusuku, 2022), así pues, periodistas del canal Willax TV, difundieron información al respecto. Asimismo, toda esta campaña estaba fundamentada básicamente por el ideario de Perú Libre, su tendencia política y sobre todo la promesa electoral de cambiar la constitución mediante una asamblea constituyente.

Ahora bien, con respecto a la candidata Keiko Fujimori, cabe mencionar que arrastra un rechazo desde que inició su carrera política, sumado al hecho de ser la heredera de los activos y pasivos del gobierno de su padre Alberto Fujimori, y también a que una cantidad de los electores son personas que no vivieron durante el periodo de gobierno de su padre (1990 – 2000) por lo tanto, la narrativa oficial de los colegios en donde han sido formados se ha centrado en los aspectos negativos de dicho gobierno (López, 2022), a esto también hay que sumarle los casos de corrupción a los cuales ha sido acusada y, al momento de redacción del presente artículo, siguen en investigación.

Bajo este marco se realizó la segunda vuelta en donde ganó Pedro Castillo con estrecho margen, pero sobre todo gracias al voto andino y rural (Miyagusuku, 2022); sin embargo, la polarización, anteriormente mencionada, fue una constante durante su corto periodo como presidente.

Desde el inicio, su gobierno estuvo acompañado por la polémica, ya que se le acusó al Jurado Nacional de Elecciones de fraude a favor de Perú Libre (BBC News Mundo, 2021), sin embargo, esta acusación no prosperó y por lo tanto no fue obstáculo para su proclamación como presidente de la república (El Peruano, 2023).

El desarrollo del gobierno tampoco estuvo exento de diversas acusaciones de corrupción, uno de los primeros casos fue dado a conocer en un reportaje de un programa dominical el 28 de noviembre del 2021 en donde se apreciaba que el mandatario recibía en forma clandestina y sin registro público a diversas personalidades, entre ellas políticos y empresarios, en un inmueble en el pasaje Sarratea en el distrito de Breña (Infobae, 2023); este acto constituye un delito en el Perú.

La relación del expresidente con el congreso fue bastante áspera, por parte de esta institución, se discutieron tres mociones de vacancia, siendo la tercera la que puso fin a su gobierno; sin embargo, Pedro Castillo fue acusado por la compra de congresistas, los denominados “niños” por el propio mandatario, a quienes habría dado sobornos a fin de que lo blindaran en el congreso en los procesos de acusación y/o vacancia (Andina - Agencia Peruana de Noticias, 2023).

Otro de los casos de corrupción más sonados fue el de los ascensos irregulares en las FFAA, el mismo que se hizo conocido en noviembre del 2021 cuando el ex comandante del ejército José Vizcarra Álvarez denunció a un medio de comunicación el haber sido presionado por personas de confianza de Pedro Castillo para favorecer a dos coroneles allegados al presidente de aquel entonces, quienes no cumplían con los requisitos para los ascensos en cuestión; esto motivó una denuncia constitucional por parte de la fiscalía de la nación (Chillitupa, 2023).

En cuanto al aspecto económico, el gobierno de Pedro Castillo representó para el Perú un incremento de la pobreza a 27.5%, lo que equivale a que más de 600 mil peruanos pasaron a dicha condición, esto debido, entre otros temas a la hostilización hacia la inversión privada (Tuesta, 2023). Por otro lado, el Centro de Investigación de Economía y Negocios Globales CIEN-ADEX estimó que el Perú perdió 11 021 millones de soles (equivalentes a 2,888 millones de dólares) por menor crecimiento económico, así como corrupción, inestabilidad económica, ineficiencia en todos los aspectos etc., lo que perjudicó en gran medida a los peruanos (Gestión, 2023).

El 7 de diciembre del 2022, Pedro Castillo Terrones da un golpe de estado sin el apoyo de las fuerzas armadas, lo que resulta en un fracaso, costándole así la vacancia y su posterior detención (Morales, 2022).

3. Metodología

La investigación fue del tipo no experimental (Hernández y otros, 2014) ya que no se aplicó ningún estímulo a la muestra, sino que se analizó en su estado natural. El universo comprende a todos los discursos oficiales del ex presidente del Perú, Pedro Castillo Terrones los cuales están comprendidos dentro del periodo 28 de julio del 2021 al 6 de diciembre del 2022, los mismos que son de dominio público y, al momento de la redacción del presente artículo, se pueden descargar desde el enlace <https://www.gob.pe/mensajepresidencial>, por lo tanto, no fue necesario tomar ninguna muestra, ya que se pudo contar con todos los discursos oficiales, los cuales ascienden a 12.

Es importante mencionar que, tanto el periodo elegido como la cantidad de discursos procesados en esta investigación, responden al hecho de que, primero son versiones oficiales, segundo son todos los que están disponibles para consulta y tercero son los únicos que ha dado oficialmente el expresidente. A continuación, la tabla 1 muestra un resumen para entender mejor el contexto en que fueron dados.

Tabla 1. Contenido principal resumido de cada discurso

N°	Fecha	Contenido principal resumido
1	28/7/2021	Asunción al poder, explicación de sus principales medidas de gobierno, así como el anuncio de una futura convocatoria a una asamblea constituyente
2	6/9/2021	Anuncio de varias medidas sociales como por ejemplo bonos familiares, también incluye medidas económicas y temas relacionados a la vacunación contra el COVID entre otros
3	6/10/2021	Acepta la renuncia del primer ministro Guido Bellido Ugarte, afirma su compromiso con la inversión privada con responsabilidad social y sin corrupción y recalca su compromiso con la gobernabilidad
4	29/9/2021	Manifiesta que un grupo de poder quiere vacarlo como presidente, también apela a sus raíces andinas para indicar que dicho grupo de poder no lo acepta como presidente, se deslinda de la corrupción y afirma que sigue trabajando en las prioridades del país.
5	4/2/2022	Menciona la necesidad de la unidad y al mismo tiempo indica que el Perú no quiere más confrontaciones, por otro lado, menciona también que el congreso le ha denegado al primer ministro exponer la política general de gobierno, también anuncia la recomposición de su gabinete ministerial, también solicita al congreso que dé prioridad a los proyectos presentados por el ejecutivo y finaliza afirmando nuevamente que tiene un compromiso firme con el país
6	27/2/2022	Rechaza las declaraciones de un aspirante a colaborador eficaz que lo vincula con un caso de corrupción, asimismo, nuevamente hace mención de que un grupo de poder minoritario no lo quiere como presidente y solicita a la fiscalía continúe con las investigaciones
7	4/4/2022	Menciona que su gobierno ha tomado las medidas necesarias para contrarrestar la crisis económica del país en aquel entonces, además indica las medidas a tomar para controlar las protestas que se iniciaban en contra de su gobierno
8	28/7/2022	Mensaje a la nación por fiestas patrias en donde explica lo avanzado en su gobierno, así como las políticas a seguir
9	9/8/2022	Mensaje a propósito del allanamiento realizado a palacio de gobierno y la casa presidencial por el poder judicial, el exmandatario menciona que se trata de un complot de un grupo de poder que busca subvertir el orden democrático
10	19/10/2022	En este mensaje el expresidente menciona que las graves acusaciones en su contra, así como la denuncia constitucional presentada por la fiscalía de la nación, son parte de un plan orquestado por un grupo de lo que él denomina "neo golpistas", que representan a los grupos de poder que no aceptan que una persona del pueblo haya llegado a la presidencia
11	24/11/2022	El expresidente Pedro Castillo menciona que el congreso le ha negado la confianza a su gabinete y con ello ha perjudicado al pueblo peruano, ya que no se va a poder ejecutar las medidas que dicho gabinete planteaba. Termina solicitando al congreso que respete el estado de derecho
12	6/12/2022	Este mensaje es dado a propósito de la tercera moción de vacancia, en este caso el exmandatario menciona que no es corrupto, que no le ha robado al país y que ejercerá su derecho a la defensa ante el congreso

Fuente: elaboración propia con información oficial del estado peruano en <https://www.gob.pe/mensajepresidencial>

Para lograr el objetivo planteado para esta investigación, es decir, determinar cuál es el mensaje detrás de los discursos oficiales del expresidente Pedro Castillo Terrones, se hizo uso de la minería de texto y de sentimientos; de tal forma que, mediante los resultados de ambas técnicas aplicadas a los discursos del exmandatario, se pueda extraer información cuantitativa que aporte al logro del objetivo anteriormente mencionado.

Los datos fueron procesados en lenguaje R, mediante la interfaz RStudio; para la minería de texto se utilizaron las siguientes librerías: dplyr, NLP, tm, SnowballC, wordcloud, magrittr; para la minería de sentimientos se utilizó: readr, NLP, tm, syuzhet, ggplot2; además para este análisis se utilizó el lexicón NRC, cuya fuente original, al momento de preparar el artículo, se ubica en el siguiente enlace: <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>, asimismo, este lexicón viene incorporado en la librería syuzhet (Jockers, 2023); Cabe mencionar que NRC clasifica a los sentimientos en dos clases: positivos y negativos, además clasifica a las emociones en: angustia, anticipación, disgusto, temor, alegría, tristeza, sorpresa y confianza, es decir las 8 emociones básicas Plutchik (Mohammad, 2021), así también es importante indicar

que es el único de los lexicones que tiene una versión en español y que puede ser invocada desde la librería antes mencionada, además la versión en español del lexicón tiene la limitación de ser una traducción de Google (Mohammad, s.f.).

Para la minería de texto, las etapas fueron las siguientes:

- a. **Preparación del archivo fuente:** los 12 discursos una vez descargados, fueron unidos en un solo archivo Word, el cual luego de una depuración mínima fue guardado como archivo de extensión txt.
- b. **Preprocesamiento del texto:** una vez listo el archivo en extensión txt, se procedió a la limpieza del texto y posterior generación de tokens. Es importante mencionar que, debido a la limpieza de palabras, algunas pueden haber sido cortadas en una letra como por ejemplo país que figura como país.
- c. **Descubrimiento y visualización:** se generó la nube de palabras, los diagramas de barras y las correlaciones, para estas últimas utiliza la correlación de Pearson dentro del comando `findAssocs`.
- d. **Análisis:** se analizaron los resultados obteniéndose información al respecto.

Las etapas de la minería de sentimiento fueron las siguientes:

- a. **Exploración del *dataset*:** en esta etapa se utilizó el archivo previamente preparado y guardado en txt utilizado en la minería de texto.
- b. **Pre-procesamiento:** se procedió a la limpieza del texto, se cargó la librería correspondiente la misma que contiene el lexicón NRC y se generaron los tokens.
- c. **Exploración y visualización:** se graficó el diagrama de barras correspondiente.
- d. **Análisis de sentimientos:** se hizo la lectura de la información generada para posteriormente extraer y analizar la información que de ella se desprendía.

Para la preparación y limpieza de texto se siguió el siguiente procedimiento en el archivo utilizado en la minería de texto:

- e. Se eliminaron los saltos de línea y tabulaciones.
- f. Se convirtió todo el texto a minúsculas.
- g. Se eliminaron los signos de puntuación.
- h. Se eliminaron los números.
- i. Se eliminaron los espacios en blanco duplicados, es decir, varios espacios en blanco seguidos se contraen en un solo espacio en blanco.
- j. Se eliminaron las *stopwords*.
- k. Se eliminaron algunas palabras puntuales que aún persistían, a pesar de la aplicación de los pasos anteriormente mencionados.
- l. Se convirtió el texto a un vector de tokens de palabras.

4. Análisis de resultados

Seguidamente, se presentan los resultados, separados por tipo de análisis realizado (minería de texto y minería de sentimientos con el lexicón NRC):

4.1. Minería de texto

A continuación, la figura 1 muestra la nube de palabras generada.

Figura 1. Nube de palabras de los discursos oficiales de Pedro Castillo Terrones

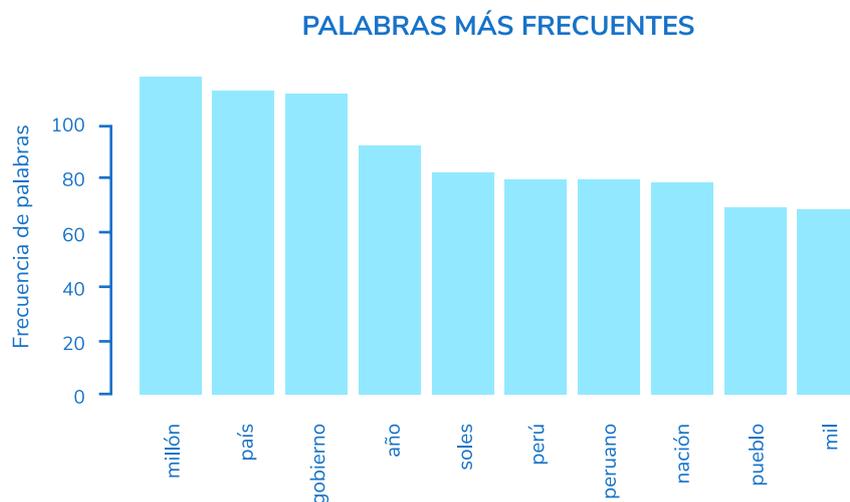


Fuente: elaboración propia.

En esta primera exploración de datos se puede observar que las palabras que más predominaron en los discursos del expresidente son: millón, gobierno, país; seguidos de: pueblo, nación, soles, Perú, peruano entre otros.

Asimismo, se ordenó el vector de palabras según frecuencia, a continuación, la figura 2 muestra el diagrama de barras ordenado de mayor a menor frecuencia del vector anteriormente mencionado.

Figura 2. Diagrama de barras de las palabras en los discursos oficiales de Pedro Castillo Terrones



Fuente: elaboración propia.

En la figura 2 se confirma la tendencia de repetición de palabras mostrada en la figura 1, en este caso se ha considerado las 10 palabras con mayor frecuencia en los discursos, de tal forma que se pueda observar gráficamente su predominancia. Otro punto importante para analizar es la correlación entre las palabras, de tal forma que se pueda observar si existe alguna asociación entre ellas, en ese sentido, la tabla 2 muestra las correlaciones superiores a 0.25 para las tres palabras que tiene mayor frecuencia en la figura 2, dicho análisis fue realizado mediante RStudio.

Tabla 2. Correlaciones de las tres palabras con mayor frecuencia

Millón		País		Gobierno	
Palabra	Correlación	Palabra	Correlación	Palabra	Correlación
soles	0.76	construir	0.34	continuará	0.31
mil	0.62	justo	0.33	existían	0.31
inversión	0.38	cambio	0.31	expectativa	0.31
asegurar	0.36	ejemplo	0.3	modernización	0.31
“bono	0.35	golpe	0.29	regulación	0.31
alimentario”	0.35	voto	0.28	agua	0.27
culminar	0.35	arriba	0.27	saneamiento	0.27
destinando	0.35	próspero	0.27	atención	0.26
distribuyendo	0.35	compromiso	0.26	necesidad	0.26
extraordinario	0.35	elegido	0.26	vigilante	0.26
implementamos	0.35	inestabilidad	0.25		
otorgaremos	0.35				
monto	0.33				
apoyo	0.32				
ejecución	0.32				
banca	0.31				
financiera	0.31				
urgencia	0.31				
yanapay	0.31				
programado	0.29				
brecha	0.28				
disminuir	0.28				
glp	0.28				
incentivar	0.28				
paquete	0.28				
aceleración	0.27				
autorizando	0.27				
contigo	0.27				
dólares	0.27				
estabilización	0.27				
pago	0.25				

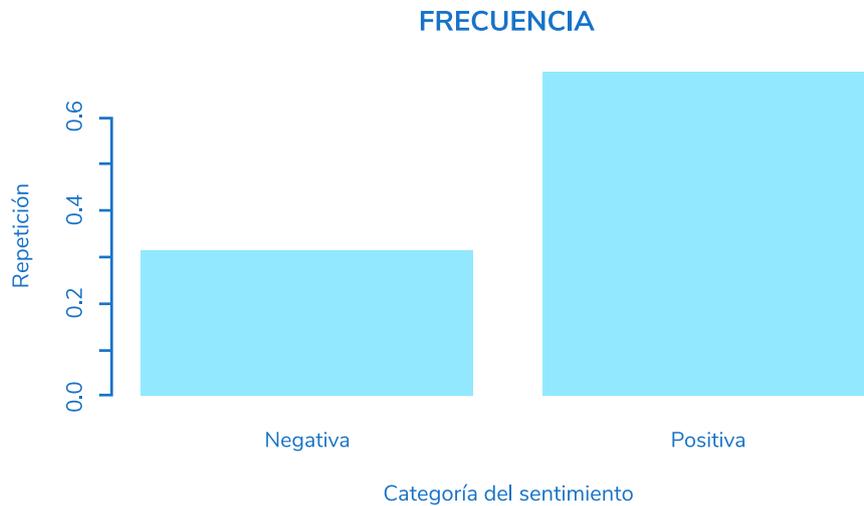
Fuente: elaboración propia.

En la tabla anterior, se analiza la correlación de Pearson de cada una de las tres palabras con mayor frecuencia mostradas en la figura 2, con el resto de las palabras del corpus y se consideran aquellas con un resultado superior a 0.25, por ejemplo: la primera columna, muestra la correlación de la palabra millón con las palabras: soles, mil, inversión etc. Así pues, se observa que solo dos palabras tienen una correlación fuerte, sin embargo, aunque con una correlación débil, existen diversas asociaciones de palabras que explican en parte el sentido de los discursos de Pedro Castillo Terrones, dichas asociaciones serán posteriormente analizadas en la siguiente sección del artículo.

4.2. Minería de sentimiento

En este primer análisis de sentimientos, se clasificó el vector de palabras en positivo o negativo, en ese marco, la figura 3 muestra una vista general de la frecuencia de las palabras con sentimientos positivos versus negativos.

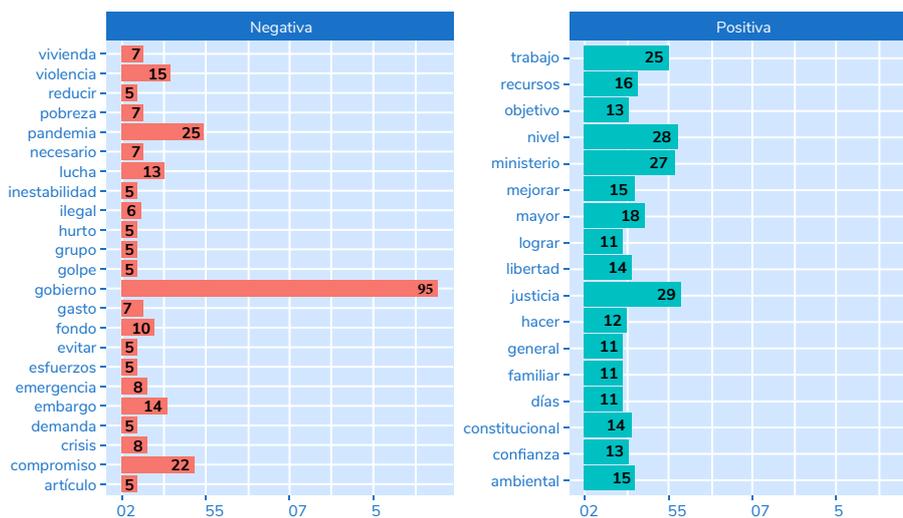
Figura 3. Diagrama de barras de la frecuencia de repeticiones tanto positivas como negativas



Fuente: elaboración propia.

Como se puede apreciar, aproximadamente un 70% corresponde a la clasificación positiva, es decir, los discursos del expresidente tienen en su mayor parte un componente de sentimientos positivos de acuerdo con el lexicón utilizado, sin embargo, ¿Qué palabras componen esta carga sentimental?, para conocer ello, la figura 4 muestra los resultados de las palabras con mayor incidencia en los sentimientos tanto positivos como negativos

Figura 4. Diagrama de barras de la clasificación de frecuencias según lexicón NRC

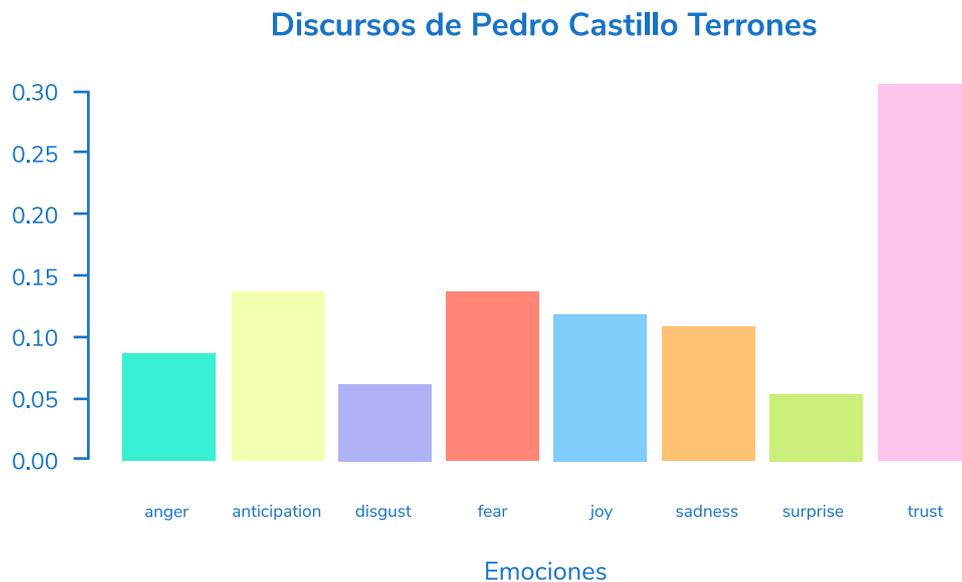


Fuente: elaboración propia.

Como se puede apreciar en la figura anterior, en la clasificación de sentimiento negativo, la palabra gobierno es la que de lejos tiene la mayor frecuencia, seguido por pandemia, compromiso, entre otros, en cuanto a la clasificación positiva, se puede observar que en primer lugar figura la palabra justicia, seguida de nivel, ministerio, etc.

A continuación, la figura 5 muestra la clasificación de sentimientos considerando las opciones que brinda el lexicon NRC.

Figura 5. Diagrama de barras de la clasificación según el lexicon NRC

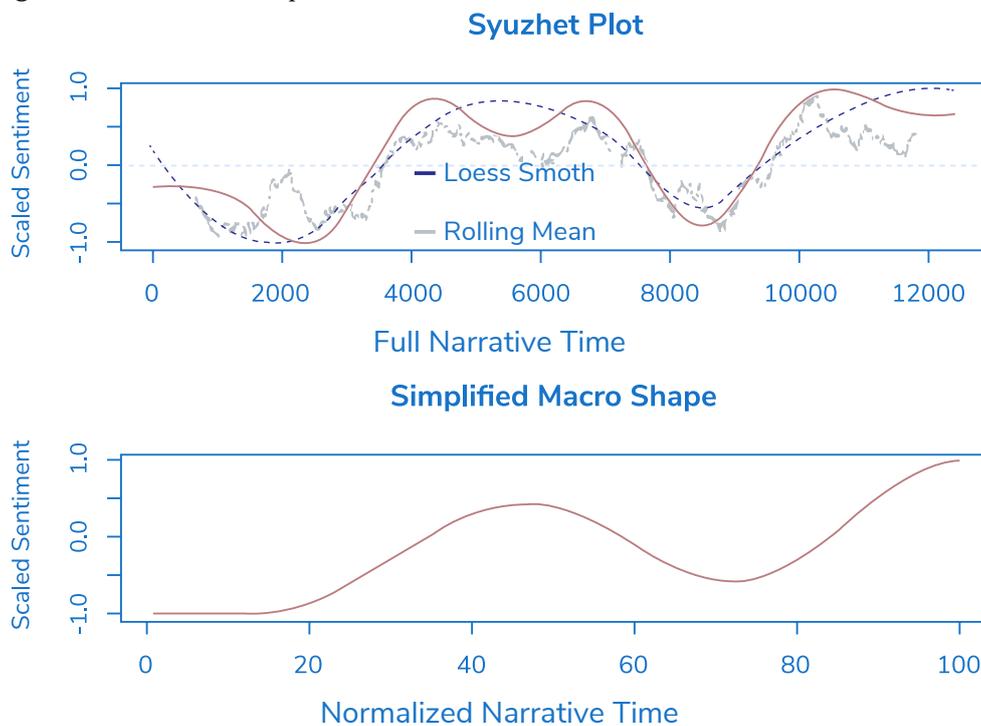


Fuente: elaboración propia.

De la figura 5 se desprende que las palabras clasificadas en la categoría confianza (*trust*) tiene una notoria predominancia sobre el resto.

Finalmente, en la figura 6 se puede observar la evolución temporal de los sentimientos en los discursos estudiados.

Figura 6. Evolución temporal de los sentimientos en los discursos de Pedro Castillo



Fuente: elaboración propia.

5. Discusión y conclusiones

5.1. Discusión de resultados

Considerando los resultados de la figura 2, a continuación, se va a analizar las tres primeras palabras: millón, país y gobierno, las cuales figuran en los tres primeros lugares; para ello se ha revisado el contexto de cada una de estas palabras dentro de los discursos analizados, las cuales han sido complementadas con las correlaciones mostradas en la tabla 2, así pues, se tiene lo siguiente:

Para la palabra “millón”, en los discursos, se observa básicamente los siguientes enfoques: primeramente, orientado describir la cantidad de beneficiarios de diversos programas, es decir, diversos colectivos de personas que han recibido algún beneficio de un determinado programa social, como segundo caso se observa que se enfocan en la descripción de los beneficios recibidos, ya sea de forma material o monetaria, esto último también se aprecia en la tabla 2.

El término “país”, entre los principales contextos dentro de los discursos, se ha verificado que está relacionado a territorio, es decir, al ámbito de impacto de las medidas planteadas o de llegada del gobierno, también está asociado a determinado conjunto de personas que demandan algo, o que tiene alguna necesidad postergada, por otro lado, también aparece dentro del contexto de vulnerabilidad y finalmente, dentro de los temas relacionados a recursos de diversos tipos, entre ellos naturales; en la tabla 2 se complementan los enfoques mostrados anteriormente, por ejemplo con el término justo que en la mayoría de los contextos de los discursos se refiere a reivindicaciones pendientes. Finalmente, la palabra gobierno, se analiza con más detalle en la discusión de los resultados de la minería de sentimientos.

Los aspectos anteriormente mencionados, tienen relación con la concepción general del rol del estado que propugna el partido Perú Libre y que se puede observar en el plan de gobierno con el que postularon a la presidencia a Pedro Castillo Terrones, es decir, un estado fuerte y presente en la mayor parte de las actividades del país, especialmente las económicas; esto se puede apreciar en su propuesta titulada economía popular con mercados, en donde el estado asume entre otros roles el de regulador de mercado, redistribuidor de la riqueza, industrializador etc. (Cerrón, 2020).

Por lo tanto, los discursos del expresidente tienen fuertes componentes asistencialistas, no fomenta el empleo, más bien la asistencia del gobierno (al menos en el discurso), así pues, no se inscriben dentro de un enfoque progresista, en el sentido de superación del ser humano, sino más de seguir dependiendo del estado.

Con respecto a la minería de sentimientos se ha observado, que en los discursos oficiales del Pedro Castillo predominan las palabras con sentimientos positivos (ver figura 3), además dentro de las palabras positivas que más se repiten, tal y como se ha mencionado anteriormente, se va a analizar las palabras justicia y nivel, ya que tienen aspectos interesantes de mencionar.

Haciendo una revisión ambas palabras en todos los discursos procesados, se puede observar que justicia obedece más que todo a la estrategia de presentar en cada mensaje, las diferentes postergaciones que viene sufriendo el pueblo peruano, sobre todo el andino, de tal forma que pueda anteponerlo ante las diversas acusaciones que se iban presentando, asimismo también con la intención de presentarse como el profesor rural que llegó al poder representando al pueblo y al cual una clase dominante busca destituirlo.

En cuanto al término nivel, esta palabra ha sido usada en distintos contextos, pero principalmente en nivel educacional, internacional, atención médica y nivel de vida, presentándose medidas que, según el expresidente, deberían mejorar estos niveles, sin embargo, se repite la estrategia anteriormente mencionada, es decir, dar la idea de que mientras el ejecutivo propone medidas para mejorar la vida de los peruanos, el poder judicial, congreso y en líneas generales los grupos de poder solo buscan destituirlo por su origen rural.

Por otro lado, dentro de las palabras con sentimiento negativo, “gobierno” es la que de lejos predomina, este hecho guarda relación con lo mencionado anteriormente en el análisis de los resultados de la minería de texto, es decir, los discursos de Pedro Castillo se circunscriben dentro de la ideología del partido que lo llevó al poder. Revisando en los discursos recopilados, la palabra gobierno (en la mayoría de los casos) está asociada a la persona del expresidente y es utilizada para indicar que es lo que va a hacer o qué medida va a tomar, en otras palabras, se busca asociar a Pedro Castillo con el poder, es decir, el que hace. También se observa su uso para referirse a coordinaciones con otras estructuras de poder formal, como por ejemplo gobierno locales, regionales etc.

Un aspecto también interesante es el mostrado por la curva normalizada de la figura 6 (curva Simplified Macro Shape) en donde se puede apreciar la evolución temporal de los sentimientos desde lo negativo (-1) hasta lo positivo (1), pasando por el neutro (0), ahí pues se observa que los discursos del expresidente comienzan con sentimientos negativos y luego comienzan a subir hacia lo positivo para luego descender y posteriormente volver a subir, es decir, siguen un vaivén entre lo positivo y negativo, sin embargo, observando nuevamente la curva generada, se puede apreciar que tiene una pendiente ligeramente positiva, en donde la segunda depresión tiene una carga menos negativa que la primera, algo parecido ocurre con la segunda cresta en donde la carga positiva es mayor que la primera.

Cabe mencionar, que la investigación se limita a los discursos oficiales de Pedro Castillo Terrones, así como al lexicón NRC utilizado, pero ¿Qué hay con los comentarios en redes sociales del expresidente? ¿mantendrán alguna relación con los discursos oficiales?, asimismo, ¿existe relación entre el mensaje detrás de los discursos y el resultado tanto económico como de la lucha anticorrupción tan prometido en su gobierno? Con respecto a la primera interrogante, sería conveniente expandir esta investigación a dichas redes sociales, en especial X y con respecto a la segunda interrogante, habría que continuar la investigación comparando los resultados económicos con los resultados de la minería de datos aplicada en este caso, sin embargo, con respecto a la corrupción, las diversas acusaciones, algunas de ellas mencionadas en el acápite del contexto, estarían demostrando que no siempre existe relación entre lo dicho y lo hecho.

6. Conclusiones

En congruencia con el objetivo de la investigación y luego del análisis de los resultados mediante la minería de texto, se ha observado que el mensaje detrás de los discursos oficiales de Pedro Castillo Terrones es la de un estado asistencialista.

También se observa, como fruto del análisis anteriormente mencionado, que los discursos del expresidente no están orientados al fomento de la iniciativa emprendedora de la población, sino más bien a la recepción por parte del estado de los recursos, en especial el económico, es decir, se fomenta la dependencia del poblador del estado; ambas conclusiones mencionadas guardan relación con el ideario o programa de gobierno del partido Perú Libre, mediante el cual llegó a la presidencia.

En cuanto a los sentimientos, la clasificación positiva es la que predomina en sus discursos, en especial los que están relacionados a la confianza, asimismo, con respecto a los sentimientos negativos, aquellos relacionados al temor, tristeza y lucha son los que tiene una mayor frecuencia.

Con respecto a su evaluación temporal, se observa que sigue un continuo vaivén que comienza con sentimientos negativos, llegando a lo positivo y luego descendiendo nuevamente, sin embargo, también se aprecia que la curva presenta una pendiente ascendente, es decir, el vaivén, anteriormente mencionado, tiende a lo positivo.

Esta investigación es un primer acercamiento hacia el mensaje detrás de los discursos del expresidente, por eso sería recomendable ampliar el análisis a sus redes sociales y posteriormente comparar lo dicho con los resultados de su gestión, de tal forma que se pueda tener un panorama integral al respecto.

Asimismo, con estos resultados se abren nuevas líneas de investigación relacionada a la aplicación de minería de texto al ámbito político.

Bibliografía

- Alaminos, A. (2023). *Introducción a la minería de texto y análisis de sentimiento con R*. Limencop.
- Andina - Agencia Peruana de Noticias. (30 de Marzo de 2023). Congreso: ¿Quiénes son los congresistas acusados por el caso 'Los Niños'? <https://andina.pe/agencia/noticia-congreso-quienes-son-los-congresistas-acusados-por-caso-los-ninos-934800.aspx>
- BBC News Mundo. (10 de Junio de 2021). *Elecciones en Perú: el partido de Fujimori pide la nulidad de 200.000 votos cuando el conteo en Perú favorece a Castillo por un estrecho margen*. <https://www.bbc.com/mundo/noticias-america-latina-57423108>
- Beynon, P. (2018). *Sistemas de bases de datos*. Reverte.
- Boden, M. (2017). *Inteligencia Artificial*. Turner.
- Centro de Noticias del Congreso. (7 de Diciembre de 2022). *Portal del Congreso de la República*. Comunicaciones Congreso. <https://bit.ly/42iurGj>
- Cerrón, V. (2020). *Perú Libre: Ideario y Programa*.
- Chillitupa, R. (12 de Agosto de 2023). *Infobae*. Fiscalía presentó denuncia constitucional contra Pedro Castillo por caso ascensos irregulares en FF AA y PNP. <https://bit.ly/498dgcS>
- Cotino, L. (2022). *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*. ARANZADI / CIVITAS.
- Dorantes, G. (2023). *Internet, sociedad y poder. Democracia digital: comunicación política en la era de la hipermediación*. UNAM, Facultad de Ciencias Políticas y Sociales.
- El Comercio. (29 de Abril de 2021). *Resultados de la ONPE: Conoce aquí las cifras oficiales con miras a la segunda vuelta*. <https://bit.ly/48TffIS>
- El Peruano. (20 de Julio de 2023). *JNE oficializa la proclamación de Pedro Castillo como presidente electo*. <https://bit.ly/47UnRr1>
- Emigidio, E. (2022). *Python: Curso Intensivo Paso a Paso Sobre Cómo Elaborar Fácilmente su Primer Proyecto de Ciencia de Datos Desde Cero en Menos de 7 Días. Incluye Ejercicios Prácticos*. Ediciones del autor.
- Fernández, E. (Ed.). (2019). *XIX Foro de investigación en comunicación. La gestión de los contenidos en comunicación*. ESIC Editorial.
- García, M. (2017). *Big Data: Gestión y explotación de grandes volúmenes de datos*. Editorial UOC, S.L.
- Gestión. (4 de Enero de 2023). *Perú perdió 2,888 millones de dólares por menor crecimiento con Pedro Castillo*. *Gestión*. <https://bit.ly/4bhvH0v>
- González, D., & Ramírez, G. (2018). *Tratado de estudios organizacionales*. Universidad EAFIT.
- Guerra, M. (2023). *Ciberinteligencia de la amenaza en entornos corporativos*. Ra-Ma S.A. Editorial y Publicaciones.
- Gutiérrez, A. (2023). *Gestionar las emociones políticas*. GEDISA.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la Investigación*. McGraw-Hill.
- Infobae. (27 de Noviembre de 2021). *Pedro Castillo: "Hay grupos que no toleran que un maestro rural haya llegado a la Presidencia de la República"*. *Infobae*. <https://bit.ly/3vVyEnp>

- Infobae. (15 de 12 de 2022). *Pedro Castillo: las ocho investigaciones en su contra que le valieron a la Fiscalía para lograr su prisión preventiva*. <https://bit.ly/4bbjRFh>
- Infobae. (7 de enero de 2023). *Infobae*. La casa de Sarratea, el lugar donde se iniciaron los negocios ilícitos del gobierno de Pedro Castillo: <https://bit.ly/495IZvT>
- Jockers, M. (11 de Agosto de 2023). *Cran R-Project*. syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text. <https://cran.r-project.org/web/packages/syuzhet/index.html>
- Joyanes, L. (2020). *Big data: Análisis de grandes volúmenes de datos en organizaciones*. Macombo.
- Jurado Nacional de Elecciones. (2021). *Reporte 9: Resultados Presidenciales de Segunda Vuelta*. <https://bit.ly/3HD04Bc>
- La República. (18 de Abril de 2021). Frases de la semana: revisa las opiniones más polémicas. *La República*. <https://larepublica.pe/politica/2021/04/18/frases-de-la-semana-revisa-las-opiniones-mas-polemicas>
- López, R. (2022). Posibles explicaciones en torno al triunfo electoral de Pedro Castillo en las elecciones peruanas de 2021: un análisis desde la antropología política. *Forum*, 22, 222-240. <https://doi.org/https://doi.org/10.15446/frdcp.n22.97450>
- Merlo, C. (2022). *Data Science Electoral*. Anderson Publishing.
- Miyagusuku, P. (2022). Elecciones peruanas 2021: desinformación y modelos en pugna. *Cuadernos del Centro de Estudios en Diseño y Comunicación*. *Ensayos*(161), 91-105.
- Mohammad, S. (2021). Sentiment analysis: Automatically Detecting Valence, Emotions, and Other Affectual States from Text. En H. Meiselman, *Emotion Measurement* (págs. 323-379). Elsevier Science.
- Mohammad, S. (s.f.). *NRC Word-Emotion Association Lexicon (aka EmoLex)*. <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- Morales, M. (7 de Diciembre de 2022). *LR::Data*. La caída de Pedro Castillo: los caminos que llevaron al exmandatario a intentar dar golpe de Estado. <https://data.larepublica.pe/el-camino-que-llevo-a-pedro-castillo-a-dar-golpe-de-estado/>
- Muñoz, P. (2023). *Apaga el celular y enciende tu cerebro: Manipulación, control y destrucción del ser humano*. HarperEnfoque.
- Orozco, G., Scolari, C., Torres, G., Piscitelli, A., Flores, C., Franco, D., & Albarrán, G. (2021). *TVMorfosis 9: Inteligencia artificial en contenidos audiovisuales*. Tintable.
- Ortega, J. (2022). *Big data, machine learning y data science en python*. RA-MA S.A. Editorial y Publicaciones.
- Quiroga, D., Montoya, C., Hernández, E., & Malit, H. (2022). *Las nuevas fuentes de productividad: perspectiva en América Latina*. Fondo Editorial – Ediciones Universidad Cooperativa de Colombia.
- Rojas, K. (13 de junio de 2022). *RPubs by RStudio*. Análisis de sentimientos en un dataset de películas: <https://rpubs.com/Karenrojas/analisisdesentimientos>
- Torres, G. (2020). *El manejo de datos Aproximaciones desde los estudios de la información*. UNAM, Instituto de Investigaciones Bibliotecológicas y de la Información.
- Tuesta, D. (15 de Mayo de 2023). *Universidad de Lima - Economía*. El legado del gobierno de Pedro Castillo. <https://www.ulima.edu.pe/pregrado/economia/noticias/el-legado-del-gobierno-de-pedro-castillo>
- TVPerú. (30 de Mayo de 2016). *TVPE Noticias*. Verónica Mendoza llama a votar por PPK: TVPerú
- Universidad de Malaga. (s.f.). *¿Qué es el Text Mining y cuáles son sus aplicaciones?* Retrieved 2023, from <https://www.bigdata.uma.es/que-es-el-text-mining-y-cuales-son-sus-aplicaciones/>