

Análisis de la situación presupuestaria 2024 de las Universidades Nacionales en Argentina según los medios de comunicación de mayor alcance a partir del Procesamiento del Lenguaje Natural

Analysis of the 2024 budget of the Argentine National Universities in the mainstream media using Natural Language Processing

Gustavo NAVARRO
Universidad Nacional de la Patagonia Austral, Argentina
 <https://orcid.org/0009-0007-1395-9308>

Cita recomendada:

Navarro, G. (2024). Análisis de la situación presupuestaria 2024 de las Universidades Nacionales en Argentina según los medios de comunicación de mayor alcance a partir del Procesamiento del Lenguaje Natural. *Publicaciones de la Asociación Argentina de Humanidades Digitales*, 5, e059.
<https://doi.org/10.24215/27187470e059>

RECIBIDO: 17 de abril de 2024 **ACEPTADO:** 10 de septiembre de 2024



RESUMEN

En el ámbito del Procesamiento del Lenguaje Natural, el análisis de sentimientos puede aplicarse en campos como las políticas públicas y gestión de reputación, percepción de clientes o usuarios, detección de tendencias en redes sociales. Este artículo explora la herramienta Streamlit, una biblioteca de Python, aplicada al análisis de la situación presupuestaria de las Universidades Nacionales en Argentina en 2024, según los cuatro medios periodísticos de mayor alcance, para detectar cómo en sus argumentos manifiestan un determinado perfil ideológico. La cuestión de la *subjetividad* versus *objetividad* sigue siendo un tema teórico, político y social de importancia en las Ciencias Humanas, particularmente ante el cisma ideológico de las últimas décadas a nivel global y fundamentalmente en nuestro país. Los resultados del trabajo evidencian polaridad en las argumentaciones, moderada objetividad y alta subjetividad en la versión de los acontecimientos que dichos medios expresan.

PALABRAS CLAVE: Recuperación de la información, análisis-contenido, lenguaje natural, texto.

ABSTRACT

In Natural Language Processing, sentiment analysis can be applied in fields such as public policies and reputation management, customer or user perception, and detection of trends in social networks. This article explores the Streamlit tool, a Python library for the analysis of the budgetary situation of the National Universities in Argentina in 2024 according to the four most far-reaching media to visualize how in their arguments they manifest a certain ideological profile. The question of *subjectivity* versus *objectivity* continues to be an important theoretical, political, and social issue in the Human Sciences; particularly in the face of the ideological schism of recent decades globally and mainly in our country. The results show polarity in the arguments, low objectivity and high subjectivity in the version of events that the media expresses.

KEYWORDS: Information retrieval, Analysis-content, Natural Language, Text.

1. INTRODUCCIÓN

En el ámbito del Procesamiento del Lenguaje Natural (PLN, también conocido como NLP por sus siglas en inglés), el análisis de sentimientos juega un papel crucial en la comprensión de la opinión y del tono del texto, ya que consiste en evaluar las emociones, actitudes y valoraciones vertidas en diferentes tipos de textos por seres humanos. De este modo, el desarrollo de herramientas eficientes para realizar esta tarea es de gran importancia en campos como la evaluación de la satisfacción del usuario o cliente y la detección de tendencias en redes sociales.

Entendiendo que desde sus inicios, la prensa se presenta a sí misma como independiente, objetiva y neutral, este análisis comparativo permite visualizar cómo diferentes medios de comunicación, lejos de ser una operación imparcial, manifiestan la defensa de valores que se

corresponden con un determinado perfil ideológico.

Aunque la cuestión de la *subjetividad* versus *objetividad* no es un tema nuevo en las Ciencias Humanas, sigue siendo de gran importancia en los ámbitos teórico, político y social; y particularmente en los últimos años recobra relevancia en función de las divisiones ideológicas binarias a nivel global y fundamentalmente en nuestro país.

El análisis acerca de la polaridad en las categorías de objetividad y subjetividad de la información expresa dicho cisma ideológico, y evidencia que en su versión de los acontecimientos, los medios expresan abiertamente sus intereses.

Este artículo propone explorar la herramienta Streamlit¹, una biblioteca de Python, que fue, en este caso, utilizada para el análisis de la situación presupuestaria de las Universidades Nacionales en Argentina en 2024, según la opinión de los cuatro medios periodísticos nacionales de mayor alcance, y entendiendo que los mismos manifiestan posicionamientos y juicios encontradas al respecto.

2. EL PROCESAMIENTO DEL LENGUAJE NATURAL COMO APROXIMACIÓN A LA COMPRENSIÓN E INTERPRETACIÓN DE TEXTOS

El PLN refiere a la rama de la inteligencia artificial y la lingüística computacional que se ocupa de la interacción entre las computadoras y el lenguaje humano; utiliza el aprendizaje automático para procesar e interpretar textos y datos. Su objetivo es permitir a las computadoras comprender, interpretar y generar lenguaje humano de manera inteligente (Bird, Klein y Loper, 2009; Hutto y Gilbert, 2014).

Algunas de sus tareas y aplicaciones más comunes incluyen: a) el análisis de sentimientos: determinar la actitud emocional en un texto, en tanto positiva, negativa o neutral, b) el reconocimiento de entidades: identificar y clasificar nombres de personas, lugares, organizaciones, fechas, etc., en un texto, c) la extracción de información: obtener

¹ Disponible en: <https://streamlit.io/>.

información estructurada de textos no estructurados tales como noticias, artículos o documentos, d) la clasificación de textos: catalogar documentos o textos en categorías predefinidas como spam o no spam, noticias o deportes, etc., e) la generación de textos: crear automáticamente texto legible y coherente basado en ciertos parámetros o datos de entrada, f) la traducción automática: descifrar texto de un idioma a otro de manera automática y precisa, g) la síntesis de textos: resumir automáticamente documentos largos o textos extensos para extraer la información más relevante, h) los asistentes de voz: la interacción de voz a texto y texto a voz para proporcionar respuestas y realizar tareas basadas en comandos de voz.

Podemos señalar estas como algunas de sus principales aplicaciones a medida que la tecnología avanza y se desarrollan nuevas técnicas y algoritmos en este campo, aunque cabe aclarar que también existe una amplia gama de otras aplicaciones en áreas como la atención médica, el comercio electrónico, el análisis de demanda, etc. En este artículo nos centraremos en el análisis de sentimientos a partir de la Streamlit.

La herramienta Streamlit fue diseñada inicialmente para aplicaciones de ciencia de datos, y en este sentido, ofrece soporte para bibliotecas populares de ciencia de datos como Pandas, NumPy y Matplotlib; esto no sólo facilita el trabajo con datos, sino que permite además visualizar los resultados. Asimismo, permite crear y desplegar aplicaciones web dinámicas para compartir y dar visibilidad a proyectos de ciencia de datos utilizando únicamente Python y sin necesidad de tener un conocimiento profundo de tecnologías web. Es decir, es una biblioteca de código abierto que le permite crear y compartir fácilmente aplicaciones web, diseñada para ser simple e intuitiva, lo que le permite implementar aplicaciones rápidamente sin escribir los extensos códigos HTML, CSS o JavaScript.

2.1. Historia y contexto del desarrollo de la herramienta

Streamlit fue lanzado en 2019 por Adrien Treuille, Thiago Teixeira y

Amanda Kelly. Su objetivo era simplificar el proceso de desarrollo de aplicaciones web para científicos que trabajan con datos, permitiéndoles crear aplicaciones interactivas utilizando sólo Python. Los fundadores combinaron sus experiencias en ingeniería de software y en ciencia de datos para crear una herramienta que posibilitara convertir fácilmente sus scripts de Python en aplicaciones web interactivas. De modo que esta herramienta se desarrolló para abordar las dificultades y complejidades que enfrentaban los científicos en el análisis y la visualización de datos.

Antes de Streamlit, la creación de tales aplicaciones requería conocimientos profundos de múltiples tecnologías web como HTML, CSS, JavaScript y el uso de frameworks como Flask o Django en el caso de Python. La motivación detrás del desarrollo de Streamlit fue entonces simplificar esos procesos y permitir crear aplicaciones web interactivas utilizando solo su conocimiento de Python. Así, la herramienta se diseñó con un enfoque en webs interactivas con solo unas pocas líneas de código. Esto hizo que el desarrollo de aplicaciones web para el análisis de datos fuera más accesible, destinado a una audiencia más amplia que no tenía experiencia previa en desarrollo web.

La información sobre la polaridad de sentimientos se considera útil para la toma de decisiones en diversos contextos de políticas públicas o gestión de reputación, también en las empresas para evaluar la percepción de los clientes o usuarios sobre sus productos y servicios. De modo que desde su lanzamiento en 2019, ha crecido rápidamente y actualmente muestra más de 10 millones de vistas de aplicaciones y más de 20 mil estrellas de GitHub.

A principios de 2022, también fue adoptada por organizaciones como la NASA y la Organización Mundial de la Salud. Actualmente, se ha vuelto extremadamente popular en la comunidad de ciencia de datos debido a su enfoque minimalista y su facilidad de uso, su sintaxis sencilla y su capacidad para integrarse fácilmente con otras bibliotecas de Python.

3. METODOLOGÍA PARA EL USO DE LA APLICACIÓN EN EL CASO ELEGIDO

Como señalamos, si bien el problema de la *subjetividad* versus *objetividad* se enmarca en la reacción paradigmática de corrientes fenomenológicas, existencialistas, sociológicas (como el interaccionismo simbólico, el construccionismo social, etc.), e incluso psicológicas (con un decidido énfasis subjetivista), el análisis de sentimientos se configura como una técnica en el campo del procesamiento del PLN que consiste en identificar, extraer y cuantificar el tono emocional presente en un texto. En el presente trabajo se utilizó para determinar la actitud general del autor (ya sea positiva, negativa o neutra hacia el tema en cuestión) en representación de una línea o posicionamiento editorial.

Los datos evaluados responden a cinco medios periodísticos destacados (Página12, Perfil, El Destape, La Nación y Clarín), y el volumen de los textos analizados en cada medio fue de 6 artículos que fueron seleccionados en función de la temática en cuestión². El período de tiempo abarcado fue durante las semanas en que la temática alcanzó su máxima exposición pública (del 8 al 26 de abril de 2024), previas a la Marcha Nacional en Defensa de la Universidad Pública que tuvo lugar en

² Véase: El Gobierno anunció un acuerdo con las universidades públicas: Qué partidas presupuestarias aumentó. *Clarín*. https://www.clarin.com/sociedad/gobierno-anuncio-acuerdo-universidades-publicas-partidas-presupuestarias-aumento_0fs9zHwoumZ.html, El presupuesto universitario es el más bajo del que se tenga registro | Su reducción es del 72 por ciento, según un estudio privado. *Página 12*. <https://www.pagina12.com.ar/729109-el-presupuesto-universitario-es-el-mas-bajo-del-que-se-tenga>, Aun con el aumento anunciado, el presupuesto para las universidades es el más bajo desde que se tiene registro. *La Nación*. <https://www.lanacion.com.ar/sociedad/aun-con-el-aumento-anunciado-por-el-gobierno-el-presupuesto-para-las-universidades-seria-el-mas-bajo-nid21042024/>, Fuerte aumento de presupuesto: El Gobierno asegura que hubo «consenso» con las universidades públicas. *Perfil*. <https://www.perfil.com/noticias/politica/marcha-universitaria-el-gobierno-asegura-que-el-ego-a-un-consenso-y-dispuso-un-aumento-del-140.phtml>, El gobierno anunció un acuerdo inexistente con las universidades, ¿qué dijo el CIN? Intento para frenar la marcha del 23. *Página 12*. <https://www.pagina12.com.ar/730188-lanzan-un-acuerdo-que-aun-no-existe>, Sobre el “costo” de la universidad, su alcance y su “eficiencia”. *El Destape*. <https://www.eldestapeweb.com/opinion/sobre-el-costo-de-la-universidad-su-alcance-y-su-eficiencia--202441723490>.

todas las jurisdicciones del país el día 23 de abril de 2024.

Para ello se utilizó la biblioteca TextBlob³ que ofrece una interfaz sencilla para realizar análisis de sentimientos en texto. Cabe aclarar que si bien las bibliotecas de Python o de otros lenguajes suelen mutar permanentemente, y es muy probable que en unos años o inclusive meses los comandos, funciones y métodos utilizados ya no funcionen, lo que siempre funcionará y se podrá replicar es la lógica que el código expresa. Esta aplicación nos permitió ingresar texto manualmente, cargar archivos en varios formatos (como .txt, .doc, .docx, .pdf y .csv), y cargar múltiples páginas web.

Las bibliotecas utilizadas se incluyeron Streamlit para la creación de la interfaz de usuario, NLTK⁴ para el análisis de sentimientos, Matplotlib⁵ para la visualización de datos, Pandas⁶ para la manipulación de datos tabulares, Requests⁷ para realizar solicitudes HTTP y BeautifulSoup⁸ para el análisis de HTML, finalmente creamos una instancia del analizador de sentimientos de NLTK. El procedimiento utilizado puede dividirse en los siguientes pasos:

- i. Entrada: se crearon diferentes funciones con diferentes tareas especializadas para extraer, procesar y ordenar la información tales como *analyze_sentiment_nltk*: toma un texto como entrada y devuelve la polaridad, subjetividad y objetividad del texto utilizando el analizador de sentimientos de NLTK; *get_text_from_url*: toma una URL como entrada y devuelve el texto de la página web correspondiente; *get_domain_name*: toma una URL como entrada y devuelve el nombre del dominio.
- ii. Interfaz de usuario con Streamlit: las aplicaciones no necesitan computadoras propias, por lo cual es una gran ventaja. El código de análisis de sentimiento se deposita en la propia computadora de

³ Disponible en: <https://textblob.readthedocs.io/en/dev/>.

⁴ Disponible en: <https://www.nltk.org/>.

⁵ Disponible en: <https://matplotlib.org/>.

⁶ Disponible en: <https://pandas.pydata.org/>.

⁷ Disponible en: <https://requests.readthedocs.io/en/latest/>.

⁸ Disponible en: <https://pypi.org/project/beautifulsoup4/>.

streamlit. La interfaz es amigable, y se pueden crear formularios, gráficos, tablas⁹.

iii. Procesamiento de la opción seleccionada; se selecciona Streamlit para armar una interfaz de usuario que permita interactuar con los datos de manera efectiva e intuitiva.

iv. Análisis de sentimientos y visualización de resultados: trasladar a streamlit el código de análisis de sentimiento desarrollado en python para que este disponible en la web¹⁰.

v. Ejecución de la aplicación: se utiliza streamlit run en la línea de comandos para ejecutar la aplicación desarrollada en streamlit; es similar a ejecutar cualquier script de Python, solo se necesita usar el comando `streamlit run appname.py`, en este caso, se aplicó `streamlit run pnl.py` en la línea de comandos de cualquier editor de python.

Una vez que los textos fueron ingresado, la aplicación realizó el análisis en base a las categorías *polaridad*, *subjetividad* y *objetividad* y mostró el resultado en forma de tabla y gráfico de barras¹¹:

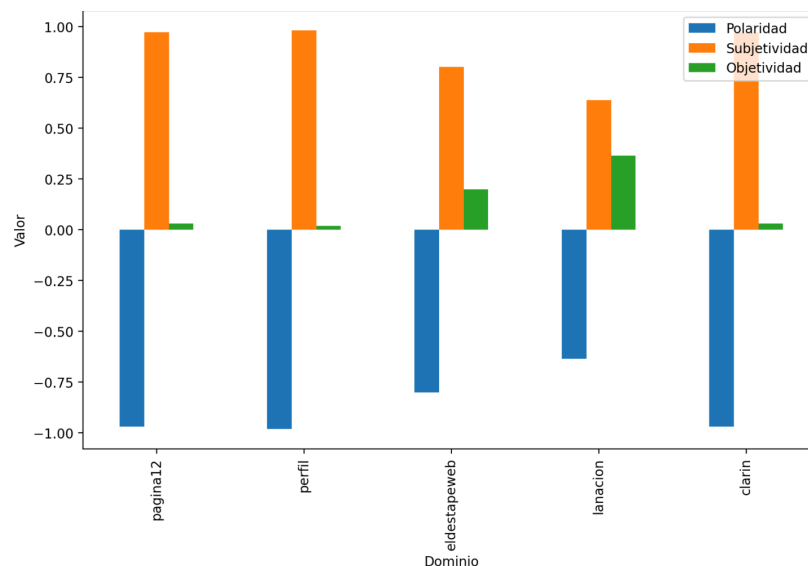


Figura 1. Gráfico de barras comparativo de polaridad, objetividad, subjetividad. Fuente: Elaboración propia.

⁹ Ver: <https://streamlit.io/>

¹⁰ Ver: <https://github.com/gusunavarro/koluel/blob/master/pnl.py>

¹¹ Ver: <https://koluel-ctdprdnbmvybw8eyzaum9.streamlit.app/>.

4. RESULTADOS DEL ANÁLISIS DE LA POLARIDAD EN LOS MEDIOS SELECCIONADOS

Para el análisis de la polaridad en los cuatro medios periodísticos ya mencionados se utilizó un volumen de 6 publicaciones referentes al presupuesto universitario en el período 8 de abril a 26 de abril de 2024. La polaridad es un aspecto fundamental del análisis de sentimientos ya que proporciona información sobre la dirección y la intensidad de lo expresado en un texto, ayuda a identificar y cuantificar emociones, juicios, actitudes, valoraciones manifiestos según sus emisores; y generalmente se mide en una escala continua que va desde negativa hasta positiva, donde valores negativos indican sentimientos negativos, valores positivos indican sentimientos positivos y valores cercanos a cero indican neutralidad o falta de sentimiento.

Del análisis realizado se desprende que los textos de los cinco periódicos analizados tienen polaridad negativa, lo que sugiere que tienden a tener una connotación negativa sobre el presupuesto universitario. Esto es central para comprender la actitud general hacia el tema en cuestión. Por otra parte, el análisis de múltiples textos, permitió comparar y contrastar los sentimientos declarados en cada uno de los medios e identificar tendencias, patrones o diferencias en su percepción.

Respecto al análisis de la subjetividad de la cobertura de noticias en los medios de comunicación, es decir, la medida en que la presentación de la información está influenciada por opiniones personales, sesgos o interpretaciones del autor o del medio, según los datos relevados, los periódicos seleccionados tienden a ser altamente subjetivos, lo que sugiere textos que contienen opiniones o puntos de vista muy marcados.

En términos de objetividad es importante partir de la consideración que si bien es sabido que ningún medio de comunicación es completamente objetivo, todos los medios están influenciados por factores como la ideología editorial, los intereses comerciales y las agendas políticas. El análisis mostró que cuatro periódicos (*Página12,*

Clarín, El Destape, Perfil) evidencian un alto grado de subjetividad; sin embargo, al menos un periódico (*La Nación*) se esforzó en presentar una gama más equilibrada de opiniones. Es decir, se desprende en uno de ellos una tendencia a ser moderadamente subjetivo, ya que incluye información meramente descriptiva que podemos catalogar como de mayor neutralidad.

Se exponen a continuación los valores relevados:

Subjetividad:

- Pagina12.com.ar: 0.9712 (altamente subjetivo)
- Perfil.com: 0.9812 (altamente subjetivo)
- Eldestapeweb.com: 0.8011 (altamente subjetivo)
- Lanacion.com.ar: 0.6369 (moderadamente subjetivo)
- Clarin.com: 0.9694 (altamente subjetivo)

Objetividad:

- Pagina12.com.ar: 0.0288 (moderadamente objetivo)
- Perfil.com: 0.0188 (moderadamente objetivo)
- Eldestapeweb.com: 0.1988 (moderadamente objetivo)
- Lanacion.com.ar: 0.3631 (moderadamente objetivo)
- Clarin.com: 0.0305 (moderadamente objetivo)

Los valores que indicarían que un resultado es altamente subjetivo/negativo o moderadamente subjetivo/negativo en un análisis de sentimientos pueden variar dependiendo de la herramienta o modelo utilizado. Sin embargo, si tomamos las siguientes pautas generales podemos considerar altamente subjetivo/negativo a:

- el porcentaje de sentimiento positivo o negativo superior a 50% (por ejemplo, 45-55),
- la puntuación de polaridad cercana a 0 (por ejemplo, -0.1 a 0.1),
- el coeficiente de variación alto (por ejemplo, > 0.5).

Y podemos considerar moderadamente subjetivo a:

- el porcentaje de sentimiento positivo o negativo entre 60-70% (por ejemplo, 55-65%),
- la puntuación de polaridad entre 0.1-0.5 (por ejemplo, 0.2-0.4),
- el coeficiente de variación moderado (por ejemplo, 0.2-0.5).

4.1. Resultado del análisis expresado en tabla y gráfico de barras

Medios	Polaridad	Subjetividad	Objetividad
pagina12.com.ar	-0.9712	0.9712	0.0288
perfil.com	-0.9812	0.9812	0.0188
eldestapeweb.com	-0.8011	0.8011	0.1988
lanacion.com.ar	-0.6369	0.6369	0.3631
clarin.com	-0.9694	0.9694	0.0305

Tabla 1. Resultados comparativos de polaridad, objetividad, subjetividad. Fuente:
Elaboración propia.

Cabe aclarar que la adecuación de las herramientas de análisis de sentimientos para el idioma español depende de varios factores, como la calidad de los datos de entrenamiento y la capacidad de la herramienta para comprender las particularidades del mismo, mucho más aún en este caso donde se evidencia el uso del dialecto rioplatense. En este sentido, fue importante elegir y evaluar las herramientas con conjuntos de datos de prueba para asegurar su eficacia.

5. CONCLUSIONES

En el marco del PLN, la aplicación de la herramienta Streamlit en el análisis de los argumentos de la situación presupuestaria de las Universidades Nacionales en Argentina en 2024 expresados en los cuatro medios de mayor alcance, nos permitió visualizar sus tendencias en términos de subjetividad y objetividad, confirmando la polarización política existente en las últimas décadas en nuestro país y en los perfiles ideológicos editoriales.

Los resultados del análisis evidenciaron polaridad en el tratamiento del tema en cuestión: moderada objetividad y alta subjetividad en la

versión de los acontecimientos que los medios de comunicación expresan. Si bien entendemos la realidad como construcción intersubjetiva de los sujetos sociales en sus diferentes manifestaciones (Zemmelman, 1992), consideramos que los contenidos vertidos por estos materializan sentidos de realidad en sus prácticas constructoras, lo que no significa mero subjetivismo o inexistencia de lo objetivo. Es decir, se advierte que los periódicos seleccionados mantienen líneas editoriales diferentes, algunos con mayor simpatía por el gobierno y otros, en clara oposición.

La aplicación de la herramienta puso de manifiesto estas tendencias como reflejo y al mismo tiempo como elemento configurador de lo social. Este trabajo e información sobre la polaridad se considera, como fue ya dicho, de gran utilidad en otros ámbitos que van desde las políticas públicas, empresas y organismos para evaluar la percepción de la ciudadanía en términos de votantes, clientes, usuarios.

REFERENCIAS BIBLIOGRÁFICAS

- Allaire, J. y Chollet, F. (2021). Keras: R Interface to 'Keras'. R package version 2.4.0. <https://CRAN.R-project.org/package=keras>
- Bird, S., Klein, E. y Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media
- Hunter, J. D. (2007). Matplotlib: A2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90-95. <http://dx.doi.org/10.1109/MCSE.2007.55>
- Hutto, C., y Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAI Conference on Web and Social Media*, 8(1), 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Zemmelman H. (1992). *Los horizontes de la razón. Uso crítico de la teoría* (pp. 131-137). Anthropos.