

Un indicador de sentimiento a partir de las NOTICIAS ECONÓMICAS

WILLY ALANYA*

Este artículo presenta un indicador de sentimiento de las noticias económicas que permite tener información de alta frecuencia (diaria) sobre las expectativas de los agentes económicos. Para su construcción, se ha procesado la recopilación periodística del Departamento de Prensa del Banco Central de Reserva del Perú (BCRP) desde inicios de 2016. El indicador propuesto muestra una alta correlación adelantada y contemporánea con los índices de difusión de expectativas del BCRP, e identifica diversos sucesos que afectaron las expectativas sobre la economía peruana en los últimos años.



* Especialista en investigación económica, Departamento de Políticas Sociales y Regionales del BCRP
willy.alanya@bcrp.gob.pe

1. INTRODUCCIÓN

En los últimos años se ha expandido el uso de métodos *machine learning* para la recolección, procesamiento y análisis de bases de datos, tanto estructuradas como sin estructuras. La subdivisión de los métodos *machine learning* que permite el análisis de información sin estructura se denomina *deep learning*. Uno de dichos métodos se basa en la lectura de textos en imágenes, es decir, en el reconocimiento natural de, por ejemplo, imágenes escaneadas de libros o capturas de pantallas de la edición digital de un periódico. Mediante el uso de estos algoritmos *deep learning*, en este artículo se desarrolla un índice para medir el sentimiento de las noticias económicas del Perú¹.

Los indicadores de sentimientos de noticia son usualmente utilizados para identificar en tiempo real tendencias en las expectativas económicas. Por ejemplo, la Reserva Federal de San Francisco² publica un indicador diario que inicia en 1980 sobre dichas tendencias. Similarmente, un trabajo de Huang y otros (2019) del Fondo Monetario Internacional (FMI) utiliza uno de estos indicadores para identificar “alertas tempranas” sobre crisis financieras en 20 países entre 1980 y 2019.

Para el caso peruano, se consideran las noticias recopiladas por el Departamento de Prensa del BCRP. La recolección de noticias comprende artículos sobre empleo, salud, inversiones, coyuntura internacional, entre otros. Al cubrir las ediciones impresas de los principales diarios de noticias económicas del país³, se asegura que la cobertura sea suficiente para inferir sobre la formación de expectativas respecto a la economía peruana.

2. METODOLOGÍA

Para la elaboración del indicador de confianza de alta frecuencia se sigue el trabajo de Algaba y otros (2023) siguiendo los siguientes pasos:

- En primer lugar, el Departamento de Prensa filtra una serie de artículos impresos y considera solo aquellos de relevancia para la economía peruana.
- En segundo lugar, las palabras de la recopilación de noticias del Departamento de Prensa se traducen del español a inglés utilizando Microsoft Office, ya que la librería R *sentometrics* para medir el sentimiento de las noticias desarrollada por Ardia y otros (2021) no cuenta con un diccionario elaborado en español.
- Posterior a la traducción, se aplica la librería *sentometrics* a cada uno de los textos obtenidos. Esta librería cuenta con tres diccionarios de palabras en idioma inglés (*lexicons*) de connotación positiva y negativa: el vocabulario General Inquirer de la Universidad de Harvard, el diccionario Henry (2008) y la terminología de Loughran y McDonald (2011) con 3 628, 189 y 2 709 palabras, respectivamente. Por ejemplo, las tres primeras columnas del Cuadro 1 muestran una lista de palabras de estos diccionarios. Se pueden encontrar términos positivos, como *accomplish*, *better* o *certainty*, y negativos, como *decline*, *difficult* o *wrong*.
- Finalmente, se emplean intensificadores y mitigadores (*valence shifters*) de las palabras. Son un total de 131 términos (*very*, *won't*, *absolutely*, etc.) que se identifican junto con las palabras de los diccionarios y que agregan o quitan valor al sentimiento de las noticias.

En específico, la fórmula para la medición del sentimiento de las recopilaciones periodísticas diarias para cada uno de los diccionarios es la siguiente:

$$s_t = \frac{1}{W} \sum_{i=0}^W val_i * palabra_i,$$

donde s_t es el valor del indicador en el día t y W es el total de palabras positivas y negativas identificadas en el diccionario. La variable *palabra* puede tomar el valor de -1 si la palabra es de connotación negativa o 1 si es de connotación positiva. Por su parte, *val* hace referencia a los *valence shifters*, cuyo valor se asigna en función del término que precede a la palabra. En caso

CUADRO 1 ■ Extracto de palabras del diccionario *Sentometrics*

Diccionario			
General Inquirer	Loughran y McDonald	Henry	Valence Shifter
<i>abrupt</i>	<i>accomplish</i>	<i>certainty</i>	<i>absolutely</i>
<i>better</i>	<i>collapse</i>	<i>increase</i>	<i>barely</i>
<i>hazard</i>	<i>danger</i>	<i>positive</i>	<i>immense</i>
<i>impair</i>	<i>instability</i>	<i>risk</i>	<i>very</i>
<i>productivity</i>	<i>profitable</i>	<i>weak</i>	<i>won't</i>

FUENTE: SENTOMETRICS.
ELABORACIÓN: PROPIA.

¹ Vega y Pinelo (2022) aplican este procedimiento para el análisis de las memorias del BCRP.

² <https://www.frbsf.org/economic-research/indicators-data/daily-news-sentiment-index/>

³ *Gestión, El Comercio, La República, Semana Económica, Exitosa, Expreso, El Peruano, Correo, Perú21, La Nación, Diario Uno*, entre otros.

de que el término intensifique el valor de la palabra (por ejemplo, “muy”, “absolutamente”, etc.), *val* toma el valor de 1,8. Sin embargo, en caso de que el término precedente mitigue la palabra, *val* toma el valor de 0,8. Finalmente, si el término cambia el sentido a la palabra de positivo a negativo, *val* toma el valor de -1.

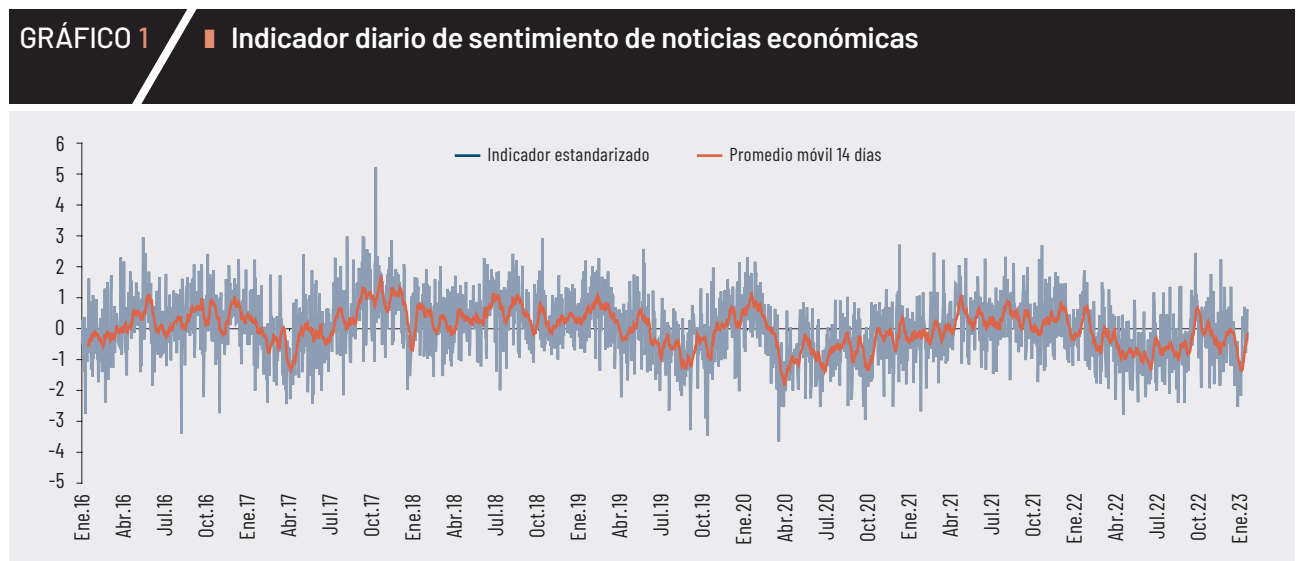
En este artículo, a diferencia de Algaba y otros (2023), se propone un indicador de noticias construido como el promedio diario de los indicadores derivados de los tres diccionarios disponibles con un aproximado de 50 artículos al día.

3. INDICADOR DE SENTIMIENTO Y CORRELACIONES DINÁMICAS

En total se han categorizado aproximadamente 57 millones de palabras reconocidas por el algoritmo en

la muestra diaria desde el 1 de enero de 2016 al 3 de enero de 2023.⁴ El Gráfico 1 muestra la evolución diaria del indicador estandarizado y la serie de medias móviles de 14 días.

La serie de alta frecuencia es volátil, pero al suavizarla se aprecian diversas tendencias alrededor de fechas importantes. Cabe precisar que la serie estandarizada diaria sigue los patrones observados en la encuesta de expectativas de la economía a 3 meses del BCRP. En la muestra, que va desde enero de 2020, se pueden identificar diversos acontecimientos relacionados a la coyuntura política peruana (proceso de vacancia, elecciones) y a la pandemia (el inicio del confinamiento y el uso de las mascarillas) que incidieron en la confianza sobre el rumbo de la economía (Gráfico 2).



FUENTE: DEPARTAMENTO DE PRENSA DEL BCRP.
ELABORACIÓN: PROPIA.



FUENTE: DEPARTAMENTO DE PRENSA DEL BCRP.
ELABORACIÓN: PROPIA.

⁴ Este estudio emplea información periodística desde el 2016; sin embargo, la información del Departamento de Prensa se encuentra disponible desde el año 2007. Actualmente se viene procesando dicha información.

CUADRO 2 ■ Correlaciones dinámicas

Variable	2016-2022	2016-2019	2020-2022
<u>Índices de Difusión BCRP</u>			
Índice de venta respecto al mes anterior (t)	0,6	0,5	0,5
Índice de órdenes de compra respecto al mes anterior (t)	0,6	0,5	0,5
Índice de la situación actual del negocio (t)	0,5	0,2	0,5
Índice de expectativas de la economía a de 3 meses (t)	0,7	0,7	0,5
Índice de expectativas del sector a de 3 meses (t)	0,6	0,6	0,5
Índice de expectativas de la demanda a de 3 meses (t)	0,6	0,5	0,5
Índice de venta respecto al mes anterior (t+1)	0,5	0,5	0,4
Índice de órdenes de compra respecto al mes anterior (t+1)	0,5	0,4	0,4
Índice de la situación actual del negocio (t+1)	0,6	0,3	0,6
Índice de expectativas de la economía a de 3 meses (t+1)	0,6	0,5	0,4
Índice de expectativas del sector a de 3 meses (t+1)	0,6	0,6	0,4
Índice de expectativas de la demanda a de 3 meses (t+1)	0,5	0,5	0,4

FUENTE: DEPARTAMENTO DE PRENSA DEL BCRP.
ELABORACIÓN: PROPIA.

En el Cuadro 2 se reporta las correlaciones dinámicas del indicador y se presenta una correlación adelantada y contemporánea importante con los índices de difusión del BCRP, en particular con las expectativas de la economía a 3 meses, donde alcanza una correlación de 0,7 al considerar toda la muestra.

A diferencia de todos estos indicadores que se elaboran a través de encuestas en determinados días del mes, el indicador de sentimiento de las noticias económicas mide diariamente la percepción de la economía. Por ejemplo, durante el proceso de vacancia de Pedro Castillo, las encuestas de expectativas del BCRP no capturaron el período de alta turbulencia que aconteció en el país, ya que se efectuaron en la segunda quincena de diciembre, cuando había un escenario político más claro. Por el contrario, tanto la caída del sentimiento como su recuperación se aprecian en la dinámica de alta frecuencia del indicador de noticias.

4. CONCLUSIONES

El indicador de alta frecuencia de noticias económicas permite observar las tendencias a tiempo real de las expectativas de los agentes económicos, ya que presenta una alta correlación contemporánea y adelanta a diversas series de expectativas del BCRP. Asimismo, el indicador evidencia diversos episodios en la economía peruana en los años de pandemia.

Futuras extensiones del indicador de sentimiento de las noticias económicas pueden comprender la creación de subíndices para diferentes variables macroeconómicas, como consumo, inversión, empleo, entre otros. Asimismo, se podría adoptar un diccionario de términos en español que podría mejorar las correlaciones dinámicas y ser un mejor predictor de las expectativas económicas. Finalmente, el presente indicador podría elevar la capacidad predictiva (*nowcasting*) de diversos modelos macroeconómicos al ser una variable de alta frecuencia.

REFERENCIAS:

- Algaba, A., Borms, S., Boudt, K., & Verbeke, B. (2023). Daily news sentiment and monthly surveys: A mixed-frequency dynamic factor model for nowcasting consumer confidence. *International Journal of Forecasting*, 39(1), 266-278. <https://doi.org/10.1016/j.ijforecast.2021.11.005>
- Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2021). The R Package *sentometrics* to Compute, Aggregate, and Predict with Textual Sentiment. *Journal of Statistical Software*, 99(2), 1-40. <https://doi.org/10.18637/jss.v099.i02>
- Henry, E. (2008). Are Investors Influenced By How Earnings Press Releases Are Written? *The Journal of Business Communication* (1973), 45(4), 363-407. <https://doi.org/10.1177/0021943608319388>
- Huang, C., Simpson, S., Ulybina, D., & Roitman, D. (2019). *News-Based Sentiment Indicators*. IMF Working Paper WP/19/273.
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66, 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Vega, M., & Pinelo, V. (2022). 100 memorias y un índice de incertidumbre económica. *Moneda*, 191, 4-9.