

Decoding Economic Insights: The Analytical Power of News Content

Pierre Delice*, David Pinto

Language and Knowledge Engineering, Faculty of Computer Sciences, BUAP, Avenida San Claudio, Blvd 14 Sur, Puebla, MEXICO.

ABSTRACT

In this article, we explore the potential of textual data as an alternative source for economic analysis. We employ a range of data extraction and labeling techniques, including domain expert, rule-based approaches and pre-trained language models to determine the polarity of textual data. By compiling a corpus of Spanish news articles from Mexico spanning 2006 to 2023 and comparing our sentiment index with economic data from official agencies, we demonstrate that news articles can serve a reliable source for economic analysis.

Keywords: Deep Learning, Text classification, MLP, Inflation, Economic indicators.

Correspondence:

Pierre Delice

Language and Knowledge Engineering,
Faculty of Computer Sciences, BUAP,
Avenida San Claudio, Blvd 14 Sur,
Puebla-72592, MEXICO.

Email: padelice@gmail.com

ORCID: 0000-0002-4170-4405

Received: 25-07-2024;

Revised: 20-09-2024;

Accepted: 14-11-2024.

INTRODUCTION

Policymakers depend on solid evidence to inform their decisions, making timely economic indicators essential for effective decision-making. The agencies tasked with compiling and disseminating these indicators face significant time constraints, both in terms of data production and making this information available to the public. As a result, there is growing interest in alternative methods for prediction. Traditional fore-casting methods, however, are fraught with uncertainty and inaccuracies, presenting substantial challenges for economists in persuading their colleagues and policymakers of the value of these predictions.

In light of these challenges, exploring the news as a potential resource for economic analysis and forecasting presents an intriguing possibility. The news media, with its real-time coverage of events, could offer immediate insights into economic trends and changes, potentially filling the gap left by the delayed release of official economic indicators. This approach could harness the vast amount of data generated daily by news outlets around the world, using advanced analytics and machine learning techniques to extract relevant economic information. By analyzing news content, economists might identify early signals of economic change, providing a more timely and possibly more accurate basis for predictions than traditional methods. This exploration

of news as a source of economic analysis and prediction could revolutionize how policymakers access and use economic data, offering a more dynamic and immediate way to inform their decisions. Utilizing one of the popular Natural Language Processing (NLP) methods, Sentiment Analysis (SA), one can extract the polarity from news articles regarding economic trends, this helps classifies text into positive, negative, or neutral categories, which are then considered for analysis. This in turn can be used for economic projection and other economic analysis.

This paper positions itself at the forefront of an interdisciplinary endeavor, anchoring its contributions within the rich and dynamic intersection of text-based sentiment analysis and economics.^[1] It explores the expansive body of research that examines the nuanced impact of linguistic tone in economic research. Through this approach, the paper contributes to the debate on two perspectives.

Firstly, our work engages with the literature strand focused on understanding the multifaceted ways in which sentiment, as conveyed through language, influences economic behaviors, market trends and policy formulations. This exploration is rooted in the premise that SA can provide a unique lens through which the psychological underpinnings of economic activities are revealed, offering fresh perspectives on consumer confidence, investor behavior and market volatility.

Secondly, the paper extends the conversation into the realm of linguistic approaches in economics, where the emphasis is placed on harnessing advanced text analysis techniques to capture and represent economic trends. In this context, our research contributes to the development of more refined models and



DOI: 10.5530/jscires.20251459

Copyright Information :

Copyright Author (s) 2025 Distributed under
Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia.[www.mstechnomedia.com]

indicators for economic analysis, pushing the boundaries of how textual data can inform economic theories and applications.

By bridging these two strands of literature, our paper not only underscores the critical role of text-based sentiment analysis in enriching economic understandings but also illustrates the transformative potential of integrating linguistic insights into economic analyses. Thus, we contribute to the ongoing evolution of both fields, proposing novel methodologies and insights that enhance our understanding of the intricate relationship between language and economic dynamics.

In particular, contributions can be outlined as follows:

We develop a Spanish corpus designed to capture trends of main economic indicators.

We create a detailed, clean, tokenized and curated dataset, establishing a crucial resource for those interested in employing NLP techniques on Spanish economic corpus. Our work showcases the capability of our corpus and dataset to accurately reflect economic polarity. This is evidenced by the creation and testing of different polarity techniques applied to our dataset.

Our work significantly propels the domain of sentiment analysis forward, particularly within the economic sphere, by introducing and refining tools and methodologies tailored for the Spanish-speaking world. More than a mere academic endeavor, our research has tangible, practical implications that transcend the theoretical realm. It offers valuable insights for crafting economic policies, conducting market analysis and making informed investment decisions. In essence, our study serves as a crucial bridge, linking theoretical research with practical solutions to real-world economic challenges faced by Spanish-speaking communities.

The structure of this article is designed to guide the reader through our comprehensive research journey. Following the introduction, we present a thorough review of existing literature, focusing on the intersection of sentiment analysis and economics, thereby setting the stage for our contributions. We then elaborate on the materials and methods employed in our study, offering detailed insights into the processes of dataset construction and cleaning process, as well as the methodologies for feature extraction, labeling and evaluation. Our development of a novel sentiment index is also discussed, underscoring its potential applications and benefits. The article culminates in a discussion that not only highlights our findings but also opens avenues for future research, emphasizing the ongoing need for innovation and exploration in the field.

Related work

The field at the crossroads of economics and sentiment analysis has seen a burgeoning body of literature. As advancements continue within the NLP community, the connection between

NLP and economics becomes increasingly evident. A fundamental question that early research endeavors to address is to what extent the tone of language can serve as a predictor of economic analysis? Tetlock *et al.* (2007) embarked on an investigation to determine if a straightforward quantitative assessment of language could forecast accounting earnings and stock returns of individual firms.^[2]

Building on Tetlock's pioneering work, a significant study by Loughran and McDonald (2011) emerged, highlighting the importance of word choice in financial analysis. They introduced a financial dictionary specifically designed to overcome the limitations of generic word lists from other fields, with the aim of more accurately capturing financial sentiment.^[3]

Despite the passage of time, many studies continue to depend on a lexicon-based approach for text classification,^[4,5] although this method faces challenges due to the polysemous nature of words—that is, words having multiple meanings or senses. This is considered as a main limitation of lexical-based for many tasks in NLP.

Consequently, alternative methods of text analysis have gained traction for offering a more nuanced understanding. Techniques such as topic modeling, word embeddings and the use of pre-trained language models have proven to be more effective in capturing the complexities of language as it relates to economic phenomena.

As mentioned above much of the application of NLP lie on finance and business and^[6] considers that financial language is inherently complex since financial terms refer to an underlying social, economic and legal context. Additionally, he stated that the type of language used within this domain is highly dependent on the context since one word or expression may have either positive or negative connotations depending on the context where it is used (stock market shares rise, debt rises, etc.). Maybe this explains the accuracy and relevance of financial data analysis and predictions.

Many works take advantage of this for focusing on forecasting^[7] presents a comprehensive sentiment engineering framework tailored for forecasting applications. This framework incorporates the elastic net technique for the selective processing of sparse data and employs a method for assigning weights to thousands of sentiment indicators^[8,9] have significantly elevated the performance benchmarks in this field.

Many of these studies typically employ specialized datasets, such as the Federal Open Market Committee (FOMC) reports,^[10] owing to the direct impact of their tone on financial markets, with Shah's research serving as a prime illustration. Alternatively, some research focuses on financial, corporate, or economic news to minimize the signal-to-noise ratio as much as possible.^[11] Additionally, some studies analyze transcripts from quarterly earnings conference calls of publicly traded

companies,^[12] leveraging these rich sources of information for their investigations.

Shah *et al.* (2022) have noted that, despite significant progress, many studies overlook tasks related to macroeconomics.^[13] Macroeconomics deals with the economy on a broad scale, covering vital elements like inflation, unemployment and economic growth, all of which are essential for a deep understanding of financial markets and economic policies. The absence of tasks focused on macroeconomics indicates that, although the model is proficient in financial text analysis, it might lack the capacity to process analyses demanding knowledge of wider economic indicators and trends. This limitation points to a promising direction for future enhancements, suggesting that incorporating macroeconomic analytical features could significantly improve the relevance and effectiveness of language models tailored for the financial sector.

Regarding the literature about NLP, advancements in text representation techniques such as Bag-Of-Words (BOW),^[14] Term Frequency-Inverse Document Frequency (TF-IDF),^[15] word2vec^[16] and Global Vectors for Word Representation (GloVe)^[17] have significantly enhanced the ability to capture the relationships between words, leading to improved accuracy. This progress is underpinned by the research of,^[18] which provides a concise overview of these foundational techniques that are critical for the development of Pretrained Language Models (PLMs).

The BOW method compiles all words in a corpus into an array, creating a mapping where each vector's value represents the frequency of a word based on its position in the text. Conversely, TF-IDF combines word frequency with the inverse document frequency to more effectively model textual information.

PLMs have brought about a paradigm shift in linguistics-based research. Their primary goal is to compile vast datasets capable of training a wide array of NLP tasks, subsequently making these models accessible for public use. A key feature of PLMs is their ability to understand context-aware word representations, addressing the challenges of polysemy and ambiguity in language. Among the most notable works in this area are BERT^[19] and RoBERTa,^[20] which exemplify the cutting-edge capabilities of PLMs in processing and understanding complex linguistic data.

Building on the insights from these significant contributions, including^[6,10] among others, our research delves into the comparative analysis of traditional entity representation against the backdrop of rule-based, lexicon and PLMs approaches for SA. This exploration aims to shed light on the evolving landscape of text representation and its implications for accurately capturing and analyzing sentiment in various contexts.

METHODOLOGY

In today's age of information overload, manually processing and categorizing vast volumes of text data presents a time-consuming and formidable challenge. Furthermore, the precision of manual text classification is susceptible to human variables, including fatigue and varying levels of expertise, leading to potential inconsistencies. Leveraging machine learning techniques to automate the text classification process emerges as a highly desirable solution, offering more consistent, objective and reliable outcomes.

Additionally, automating this process significantly improves the efficiency of information retrieval, effectively mitigating the issue of information overload by streamlining the search for specific information. Contrary to most of previous work where labelling

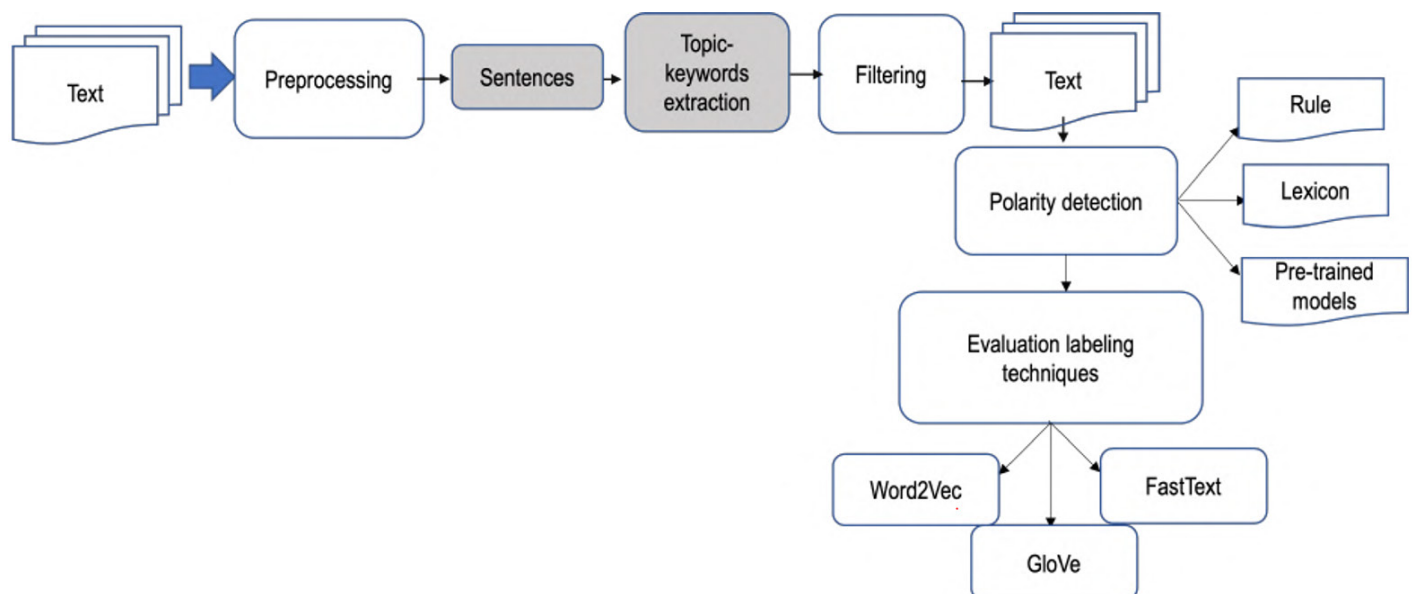


Figure 1: Methodological approach

Table 1: Lexicon Values.

Lexicon	Positive words	Negative words
ElhPolar	1,897	3,302
iSOL	2,509	5,626
ML-SentiCON	5,568	5,974
ElhPolar \cup iSOL \cup ML-SentiCON	8,146	11,959

Table 2: Rule-based classification

Panel A1	Panel B1
inflation expectation, interest rate, bank rate, fund rate, price, economic activity, inflation, employment	unemployment, growth, exchange rate, productivity, deficit, demand, job market, monetary policy
Panel A2	Panel B2
anchor, cut, subdue, decline, decrease	ease, easing, rise, rising, increase, expand

Table 3: Sentence-splitting result.

	Mean		Total	
	Sentences	Words	Sentences	Words
Before Splitting	26.40	973.47	1,650,282	60,853,461
After Splitting	1.00	39.81	312,635	12,442,406

includes human involvement, in our case, we use 3 approaches for labelling: rule, lexicon and transfer-learning (see Figure 1).

Lexicon

A lexicon-based Sentiment Analysis (SA) approach represents an unsupervised methodology that leverages a curated dictionary, wherein words are annotated with positive or negative sentiment values. This technique is grounded in the meticulous assembly of linguistic elements, such as morphemes, which are enriched with expert insights.^[21] It employs a direct lookup mechanism to ascertain the sentiment polarity of words.

In our case, we employ a collection of lexicons documented in the literature, namely ElhPolar, an English polarity lexicon that has been translated into Spanish, where any ambiguous translations manually resolved by annotators. It contains 1,897 positive words and 3,302 negative words. On the other words, iSOL which is a Spanish words indicating domain independent polarity contains 2,509 positive words and 5,626 negative words. Finally, ML-SentiCON which is a multilingual layered sentiment lex- icons at lemma level contains 5,568 positive words and 5,974 negative words. Taken the intersection of the three lexicons, we definitely compiled 8,146 positive words and 11,959 negative words (see Table 1).

Rule-based

As stated by Shah *et al.*^[10], in the realm of finance rule-based classification systems has been a standard practice. These systems typically classify information based on the detection of specific keyword combinations, notably demonstrated the efficiency of such methods by categorizing sentences as either positive or negative, depending on the mix of financial nouns and verbs within certain sentence segments.

We decide to explore those rules in our corpus in order to analyze whether extraction of financial terminology and tone is a suitable approach for classification. According to Table 2, a sentence is deemed positive if it includes terms from either panels A1 and A2 or B1 and B2. Conversely, sentences are labeled negative if their feature terms from A1 and B2 or A2 and B1. Our goal is to evaluate the rule-based method's efficiency on our datasets as a comparative measure to the deep learning approaches we plan to implement subsequently. This rule-based method is applied to test datasets derived from each dataset using an 80:20 split for training and testing.

The sentiment of a sentence or phrase is determined by first identifying the sentiment score for each word from the chosen lexicon and then adding them up to arrive at the overall sentiment.

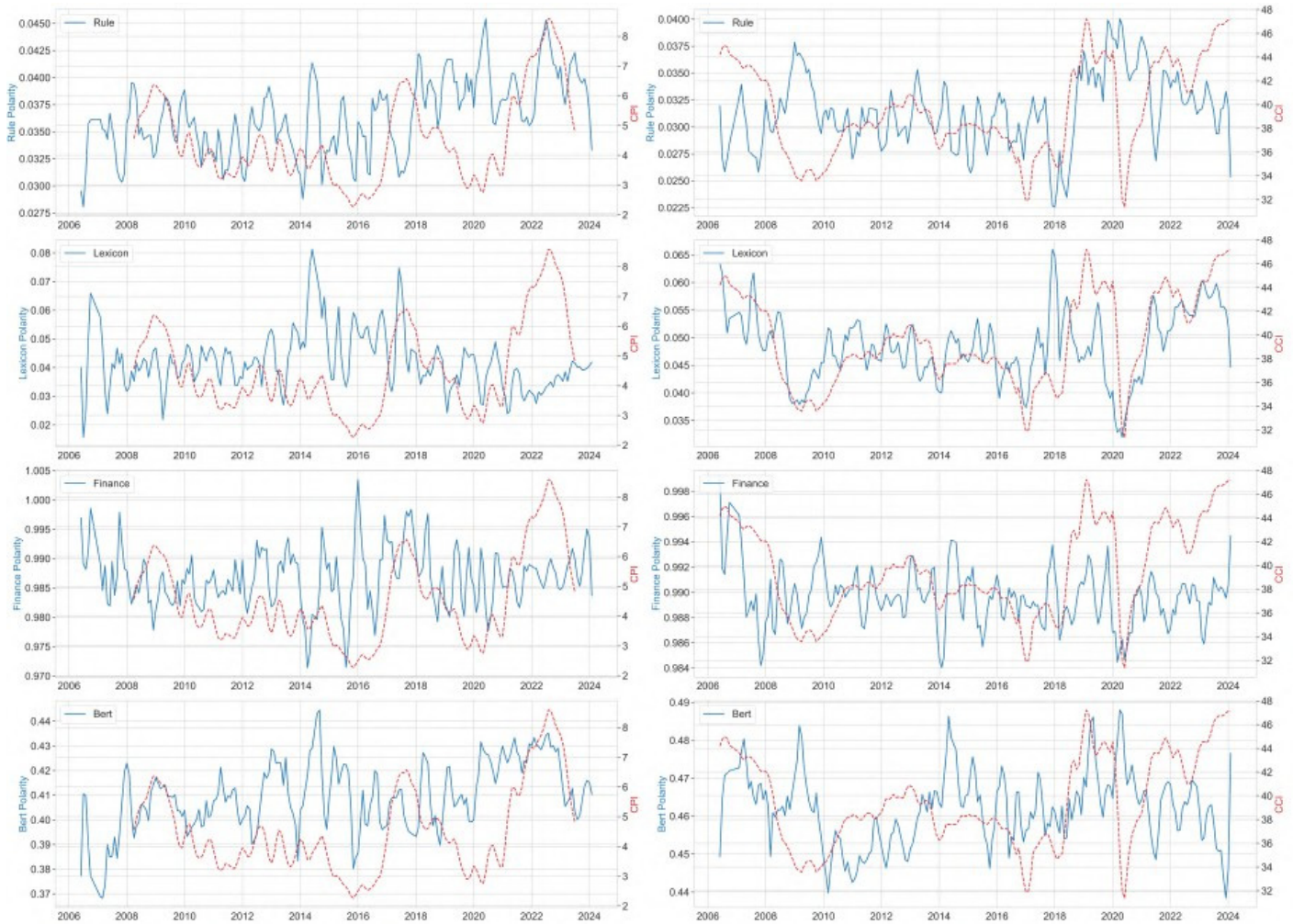


Figure 2: Comparison of Consumer Price Index vs Rule and Lexicon.

Pre-trained Language Models

Finally, we use transfer learning techniques from two Pre-trained Language Models based on Bidirectional Encoder Representations from Transformers (BERT) to label our dataset. FinBERT which is a financial SA trained on financial news article to capture prices movement in an open market.^[10,22]

On the other hand, SentiBERT captures compositional sentiment semantics and demonstrate competitive performance on phrase-level sentiment classification.

Dataset

In this section, we present the composition and source of the datasets used in our study. Our primary dataset is meticulously constructed through web scraping techniques, encompassing a comprehensive collection of raw text extracted from various news outlets.

This dataset is uniquely curated from articles published in the economic sections of these outlets, spanning a significant period from 2006 to 2023. By focusing exclusively on economic reporting,

we ensure that our dataset is rich in domain-specific language and themes, providing a robust foundation for our deep learning models to analyze and derive insights into macroeconomic indicators.

News Data

We use the repository from the Center of Documentation and Information Mtro. Jesu's Silva Herzog at the Autonomous National University of Mexico (UNAM). For the content downloading, we employed the Beautiful Soup and Selenium libraries. Following this phase, we used an array of libraries, including spaCy and NLTK, to clean and preprocess textual data.

The first collection yielded a total of 73,512 news items, spanning from May 2006 to January 2024. Our preliminary data cleaning process involved tasks such as completing the date column from raw text and implementing a double-crawling strategy from main news outlets when content was initially unavailable. These efforts resulted in a refined database comprising approximately 62,512 daily news items.

Table 4: Topics and their top words with weights.

Topic 1	Topic 2	Topic 3	Topic 4
market (2.19%)	investment (1.93%)	growth (2.94%)	prices (3.08%)
markets (1.85%)	market (1.32%)	inflation (1.67%)	index (1.26%)
interest (1.70%)	mexico (1.09%)	economy (1.33%)	inflation (1.16%)
united (1.34%)	growth (0.92%)	economic (1.04%)	dollars (1.04%)
rates (1.26%)	investments (0.76%)	mexico (0.97%)	products (1.01%)
investors (0.91%)	companies (0.70%)	activity (0.87%)	month (0.96%)
prices (0.85%)	sector (0.66%)	demand (0.86%)	annual (0.93%)
dollar (0.77%)	government (0.55%)	global (0.82%)	price (0.81%)
bank (0.76%)	if (0.44%)	economic (0.75%)	points (0.79%)
week (0.75%)	larger (0.42%)	quarter (0.75%)	increase (0.74%)
oil (0.69%)	national (0.41%)	bank (0.74%)	pesos (0.71%)
reserve (0.62%)	industry (0.40%)	countries (0.54%)	national (0.65%)
index (0.60%)	labor (0.38%)	china (0.54%)	level (0.64%)
rate (0.56%)	dollars (0.37%)	united (0.51%)	increment (0.63%)
peso (0.53%)	also (0.37%)	larger (0.50%)	last (0.58%)
inflation (0.52%)	development (0.37%)	recovery (0.50%)	mexico (0.57%)
stock market (0.50%)	president (0.36%)	global (0.47%)	months (0.55%)
American (0.47%)	spending (0.35%)	said (0.45%)	larger (0.54%)
federal (0.47%)	years (0.34%)	markets (0.44%)	food (0.53%)
shares (0.47%)	part (0.33%)	policy (0.43%)	rate (0.52%)

Models

Following the data cleaning phase, we embark on conducting a Latent Dirichlet Allocation (LDA) analysis, selecting an appropriate number of topics for each country under study. This step involves a detailed interpretation of the semantic themes that emerge from the terms associated with each topic. Our focus then narrows to identifying and collecting topics that are pertinent to inflation, economic growth and financial markets.

Within these identified topics, we meticulously examine the words, specifically compiling a list of nouns that carry significant economic relevance. This process allows us to distill the essence of the discussions surrounding key economic indicators, providing a nuanced understanding of the economic landscape as depicted through the textual data. This curated set of economically meaningful nouns serves as a cornerstone for further analysis, enabling us to gain deeper insights into the economic conditions and sentiments expressed in the corpus.

To construct the time series of the NSI, the news articles are sorted by date of publication and the data are divided into blocks of time, which could be a day. For each time period (t), we compute the sentiment index by subtracting the count of negative words from the count of positive words and then dividing by the total word count:

$$SI_t = \frac{Positive_t - Negative_t}{WordCount_t}$$

In the final stage of our analysis, we employ a classifier based on a Multi-Layer Perceptron (MLP) architecture, which is integrated with various entity distributions to assess the effectiveness of our label sentiment in capturing the overall tone present within our corpus. To enhance the robustness and accuracy of our sentiment analysis, we incorporate two additional pre-trained models specifically designed for analyzing sentiment in Spanish: BERT and Bardsai.

These models, renowned for their deep learning capabilities and nuanced understanding of language, are utilized to cross-validate the sentiment index we have developed. By leveraging the strengths of BERT, known for its powerful contextual language understanding and Bardsai, which may offer specialized features for sentiment analysis in Spanish, we aim to ensure a comprehensive and accurate assessment of the sentiment conveyed in our text data.

This multi-model approach allows us to compare and contrast the performance of our sentiment index against the benchmarks set by these pre-trained models. It also provides a deeper insight into the nuances of sentiment expression in Spanish-language text, ensuring that our analysis is thorough and sensitive to the

Table 5: Granger P-Values and Partial Correlation.

		Partial Correlation	Granger <i>p</i> -Value			
			Lag 1	Lag 2	Lag 3	Lag 4
CCI	rule	0.37	0.91	0.26	0.40	0.45
	lexicon	0.60	0.00	0.02	0.05	0.03
	bert	0.10	0.20	0.58	0.23	0.05
	finance	0.03	0.74	0.88	0.75	0.43
CPI	rule	0.10	0.13	0.21	0.37	0.03
	lexicon	-0.20	0.45	0.62	0.87	0.60
	bert	0.22	0.10	0.16	0.29	0.43
	finance	0.22	0.88	0.21	0.28	0.40

Table 6: Comparison of accuracy (%) of sentiment classification.

Model	Embeddings	Rule	Lexicon	Finance	Bert
	Word2vec	68.20	87.97	70.25	82.31
MLP	GloVe	66.94	79.73	52.56	79.04
	FastText	73.58	88.39	76.16	85.50

linguistic and cultural contexts inherent in the data. Through this process, we seek to confirm that our sentiment index accurately reflects the tones and nuances captured in our corpus, thereby validating the effectiveness of our analytical framework.

RESULTS AND ANALYSIS

In order to evaluate our rule- and lexicon-based sentiment index, we compare each index with the Consumer Price Index (CPI) and the Consumer Confidence Index (CCI) and measure whether they can capture the main economic trend. The CPI is an economic indicator that measures the average price variation of representative commodities of the country's household consumption over time. These price variations directly impact the purchasing power and well-being of consumers, making the CPI a significant and generally interesting indicator for society. While the CCI measures the level of confidence of the consumer in the markets.

Inflation is among the main economic concept discussed in news media due to its impact on household^[23] and according to the result of the topic modelling from the Table 4, words related to inflation such as: price, market, dollar contribute significantly in the 4 latent topics.

Rule-based sentiment analysis evaluates the tone of news based on how inflation affects consumer purchasing power; thus, an increase in inflation correlates with a decline in the sentiment index. This is why the first panel of Figure 2 illustrates an inverse relationship between inflation and the rule-based sentiment index, highlighting the direct impact of economic conditions on public sentiment as captured through this analytical approach.

In the second panel, the focus shifts to comparing lexicon-based sentiment with the CPI. The lexicon approach assesses the linguistic tone of the text corpus, capturing shifts in sentiment related to the economic narrative. Consequently, as inflation increases, lexicon-based analysis tends to emphasize the negative impacts of inflation. This methodology explains why, for this index, we observe a trend that aligns more closely with inflation, contrasting with the inverse trend seen in the rule-based index.

This alignment underscores the nuanced ways in which different sentiment analysis approaches can reflect economic realities.

In the Table 5 we calculate the correlation between CPI, CCI and each method. As well as we calculate the Granger test, which is a statistical hypothesis test for determining whether one-time series is useful in forecasting another. Although named as such, it is important to note that the Granger test does not determine causation in the philosophical sense. Instead, it tests whether one variable provides statistically significant information about the future values of another.

Employing a MLP with three hidden layers, each consisting of 128 neurons, our analysis reveals that the sentiment index derived from the lexicon-based approach tends to outperform other methodologies. Specifically, we observed a range of accuracy between 79.7% and 88.4% (see Table 6).

CONCLUSION

This study leverages textual data to contrast SA with numerical datasets. We employed three distinct methodologies to annotate our dataset: rule-based, lexicon-based and utilizing pre-trained models. By evaluating these labeling technique through a MLP, we observed that the lexicon-based SI yielded satisfactory result. Our

contribution is twofold: we provide a meticulously cleaned and annotated dataset in Spanish, specifically curated for economic analysis and we offer insights into the effectiveness of different sentiment labeling approaches. This enriches the resources available for economic SA, particularly for Spanish language datasets and under-scores the potential of lexicon-based methods in extracting meaningful sentiment from textual data.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

- GONZÁLEZ M, TADLE RC. Signaling and financial market impact of chile's central bank communication: A content analysis approach. *Economía* 20(2), 127-78 2020. [accessed 2024-2-15].
- Tetlock PC, Saar-Tsechansky M, Macskassy SA. More than words: quantifying language to measure firms' fundamentals. *Texas finance festival*; 2007.
- Loughran T, Mcdonald B. Textual analysis in finance. *Behavioral & Experimental Finance (Editor's Choice) eJournal*; 2020.
- Bos T, Frasinca F. Automatically building financial sentiment lexicons while accounting for negation. *Cognit Comput*. 2022;14(1):442-60. doi: 10.1007/s12559-021-09833-w.
- HASSAN TA, HOLLANDER S, LENT LV, TAHOUN A. The global impact of brexit uncertainty. *The Journal of Finance*. 2024;79(1):413-58. doi: 10.1111/jofi.13293.
- Pan R, García-Díaz JA, García-Sánchez F, Valencia-García R. Evaluation of transformer models for financial targeted sentiment analysis in Spanish. *PeerJ Comput Sci*. 2023;9:e1377. doi: 10.7717/peerj-cs.1377, PMID 37346571.
- Ardia D, Bluteau K, Boudt K. Questioning the news about economic growth: sparse forecasting using thousands of news-based sentiment values. *Eur Econ Monet Econ ej*. 2017.
- Araci D. Finbert: Financial sentiment analysis with pretrained language models. *ArXiv abs/1908.10063* 2019.
- Gössi S, Chen Z, Kim W, Bermeitinger B, Handschuh S. Finbert-fomc: fine-tuned finbert model with sentiment focus method for enhancing sentiment analysis of FOMC minutes. In: *Proceedings of the fourth ACM international conference on AI in finance*. ICAIF'23. New York; 2023. <https://doi.org/10.1145/3604237.3626843>. p. 357-64. Association for Computing Machinery. doi: 10.1145/3604237.3626843.
- Shah A, Paturi S, Chava S. Trillion dollar words: A new financial dataset, task & market analysis. In: *Annual Meeting of the Association for Computational Linguistics*; 2023. Available from: <https://api.semanticscholar.org/CorpusID:258685646>. doi: 10.18653/v1/2023.acl-long.368.
- Kalamara E, Turrell AE, Redl CE, Kapetanios G, Kapadia S. Making text count: economic forecasting using newspaper text. *SSRN Electron J*. 2020. doi: 10.2139/ssrn.3610770.
- Hassan TA, Lent L, Hollander S, Tahoun A. The global impact of brexit uncertainty. *Microeconomics Decis-Mak Risk Uncertainty ej*. 2019.
- Shah RS, Chawla K, Eidnani D, Shah A, Du W, Chava S, *et al*. When flue meets flang: benchmarks and large pretrained language model for financial domain. In: *Conference on Empirical Methods in Natural Language Processing*; 2022. Available from: <https://api.semanticscholar.org/CorpusID:253244049>. doi: 10.18653/v1/2022.emnlp-main.148.
- Zhang Y, Jin R, Zhou ZH. Understanding bag-of-words model: a statistical framework. *Int J Mach Learn Cybern*. 2010;1(1-4):43-52. doi: 10.1007/s13042-010-0001-0.
- Pino JA. Modern information retrieval. *ricardo baeza-yates y Berthier ribeiro-neto addison wesley hariow, England*; 1999. Available from: <https://api.semanticscholar.org/CorpusID:161232122>.
- Mikolov T, Chen K, Corrado GS, Dean J. Efficient estimation of word representations in vector space. In: *International Conference on Learning Representations*; 2013. Available from: <https://api.semanticscholar.org/CorpusID:5959482>.
- Pennington J, Socher R, Manning CD. Glove: global vectors for word representation. In: *Conference on Empirical Methods in Natural Language Processing*; 2014. Available from: <https://api.semanticscholar.org/CorpusID:1957433>.
- Li Q, Peng H, Li J, Xia C, Yang R, Sun L, *et al*. A survey on text classification: from traditional to deep learning. *ACM Trans Intell Syst Technol*. 2022;13(2):1-41. doi: 10.1145/3495162.
- Devlin J, Chang MW, Lee K, Toutanova K. Bert: pretraining of deep bidirectional transformers for language understanding. In: *North American Chapter of the Association for Computational Linguistics*; 2019. Available from: <https://api.semanticscholar.org/CorpusID:52967399>.
- Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, *et al*. Roberta: A robustly optimized bert pretraining approach. *ArXiv abs/1907.11692* 2019.
- Albrecht J, Ramachandran S, Winkler C. Blueprints for text analytics using python. *O'Reilly Media*; 2020. p. p.? Available from: <https://books.google.com.mx/books?id=ad0MEAAQBAJ>.
- Gorodnichenko Y, Pham T, Talavera O. The voice of monetary policy. *CEPR Monet Econ Fluctuations (Top)*. 2021.
- Larsen VH, Thorsrud LA, Zhulanova J. News-driven inflation expectations and information rigidities. *J Monet Econ*. 2021;117:507-20. doi: 10.1016/j.jmoneco.2020.03.004.

Cite this article: Delice PA, Pinto D. Decoding Economic Insights: The Analytical Power of News Content. *J Scientometric Res*. 2025;14(1):365-72.