Irving Fisher Committee on
Central Bank Statistics

IFC-Bank of Italy Workshop on "Data science in central banking: enhancing the access to and sharing of data"

17-19 October 2023

Constructing high-frequency and thematic economic sentiment indicators from online news articles: applications in the Philippine context[1]

Alan Chester Arcin, Carmelita Esclanda-Lo, Chelsea Anne Ong and Rossvern Reyes,
Bangko Sentral ng Pilipinas

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Constructing High-Frequency and Thematic Economic Sentiment Indicators from Online News Articles

Applications in the Philippine Context

Alan Chester Arcin, Carmelita Esclanda-Lo, Chelsea Anne Ong and Rossvern Reyes[1]

Abstract

The use of natural language processing (NLP) methods in informing macroeconomic policy has risen over the past decade with the increasing availability of alternative data sources. Digital news articles are high-frequency sources of information that economic agents may refer to when making decisions and policies. In this study, we create a News Sentiment Index (NSI), a high-frequency indicator that is based on online business and financial news, to infer market sentiment. Specifically, two NSIs are constructed using the dictionary and machine learning methods, with a sentiment lexicon created from genetic algorithm and Financial Bidirectional Encoder Representations from Transformers (FinBERT), a pre-trained NLP model trained in the finance domain. To examine the relevance of these indicators in capturing key global and Philippine economic events, we evaluate the relationship between the NSIs and other economic indicators through correlation analysis. Furthermore, this study expands to topic modeling using Non-negative Matrix Factorization to explore sentiments by theme. Our findings show that the NSIs track key economic events and are positively and strongly correlated with other high-frequency indicators, such as the Purchasing Managers' Index (PMI) and Philippine Stock Exchange Index (PSEi). Meanwhile, on topic modeling, sentiments from news on company earnings and trade also show strong positive correlations with the PMI, PSEi, and survey-based outlook indicators. We conclude that use of NSIs and topic modeling can potentially complement existing survey-based measures to timely gauge market sentiment.

Keywords: Sentiment Analysis, Opinion Mining, Text Analysis, Dictionary, Machine Learning

JEL Classification: C82 D83

1 This paper was prepared for the IFC and Bank of Italy Workshop on "Data Science in Central Banking" held on 17-19 October 2023 in Rome, Italy. Alan Chester Arcin (arcinat@bsp.gov.ph) is from the Department of Economic Research (DER) of the Bangko Sentral ng Pilipinas (BSP), Carmelita Esclanda-Lo (esclandacg@bsp.gov.ph), Chelsea Anne Ong (ongcs@bsp.gov.ph), and Rossvern Reyes (reyesros@bsp.gov.ph) are from the Department of Economic Statistics (DES) of the BSP. The views expressed herein are those of the authors only and do not necessarily reflect those of the BSP. The authors acknowledge Cherrie Mapa and Genna Paola Centeno of the DER-BSP for their contributions to this study. The authors also thank Senior Director Redentor Paolo M. Alegre, Jr. of the DES-BSP and Officer-in-Charge Dennis D. Lapid of the DER-BSP for their unwavering support.

Contents

1. Introduction

Sentiment of economic agents can have a considerable impact on real economic activity. For instance, consumer sentiment can influence consumption and saving decisions, while business sentiment can affect investment, hiring, and pricing decisions. As such, a measurement of market sentiment is an important component of macroeconomic surveillance.

Typically, sentiment or people's attitudes or opinions towards a particular topic are measured through surveys. The Bangko Sentral ng Pilipinas (BSP) deploys the Business Expectations Survey (BES) and Consumer Expectations Survey (CES) every quarter to gauge the economic sentiment of firms and households, respectively. However, conducting surveys can be tedious, often time-consuming, and expensive. The lags in the release of survey results imply that information that could be important for economic surveillance and policymaking may not be available on a near-real-time basis.

More recently, the accessibility of possible alternative high-frequency sources of information, such as online newspaper articles and the growing field of sentiment analysis techniques, has opened opportunities for innovative sentiment measurement. This study explores applying sentiment analysis techniques on news data to determine market sentiment and generate news-based sentiment indices (NSIs). Two indices are developed using: (1) dictionary method and (2) machine learning (ML) method.

This study adds to the increasing list of central banks (Federal Reserve Bank of San Francisco (Shapiro et Al., 2020), Bank Negara Malaysia (Chong et al., 2022), Bank of Korea (Seo, 2022), Reserve Bank of Australia (Nguyen & Cava, 2020), and Narowdy Bank Polski (Marszal, 2022)) that performed sentiment analysis of online news data as an economic surveillance tool.

Overall, this research shows that the NSIs match the expected sentiment of key economic events. These indices display the potential to supplement existing surveys and provide a timely measurement of market sentiment. This study also delves into topic modeling to identify key themes that could be useful in determining the behaviour of certain key economic indicators.

2. Data

The data used in this study consists of news articles from multiple sources on the web. Table 1 shows the news sources used in the dataset, sorted by share to total in descending order.[2] The availability of digital archives allows news articles to be retrieved as early as January 2018. We are particularly interested in the sections of economy, business, finance, and related fields (e.g., stock market, banking).

[2] The authors have coordinated with several media outlets to seek permission to scrape their news data. The following media outlets have agreed to have their news articles retrieved online: BusinessWorld, Manila Bulletin, Inquirer.NET, Manila Standard, and BusinessMirror.

Share of News Sources to Total Number of Articles Gathered

January 2018 – December 2022 Table 1

	Average number of articles per month	Share to the total number of articles gathered
BusinessWorld	453	25.9%
Manila Bulletin	520	23.8%
Manila Standard	332	18.7%
BusinessMirror	294	16.8%
Inquirer.NET	260	14.8%

Source: Authors' Estimates

3. Methodology

Text Pre-processing

Pre-processing of text was done to reduce the noise and unwanted parts of the data and subsequently improve the accuracy of models used in the study. We used the following standard NLP practices in cleaning text:

- Standardization of column headers and date format;
- Exclusion of ads and sponsored articles;
- Deletion of the *lede* (the first paragraph summarizing the article) of some news sources as these may be redundant to the content of the news article;
- Removal of extra spaces and unwanted punctuation marks, hyperlinks, and HTML tags;
- Case normalization;
- Conversion of Unicode characters to ASCII;
- Removal of stop words; and,
- Tokenization of articles.

Note that dictionaries might assign a different polarity for base words and their inflections (e.g., inflate and inflation), and stem words interfere with word matching. Hence, we did not use lemmatization and stemming.

Approaches in Sentiment Analysis

Broadly, sentiment analysis is used to determine the tone of text, whether positive, negative, or neutral. We utilized two general approaches in determining sentiment: (a) dictionary-based method and (b) ML-based method.

Dictionary-based method

The dictionary-based approach involves lexicons and a set of rules. While this word-matching method is straightforward, we acknowledged that it trades interpretability and speed with one of its obvious flaws – ignoring word context and complex word usage. We used the following well-known lexicons in literature listed in Table 2:

List of Sentiment Lexicons

Selected Lexicons with various domains — Table 2

	Description	Author
General Inquirer (GI Lexicon)	Created for general-purpose content analysis of text data	Stone et al. (1966)
Loughran-McDonald Sentiment Dictionary (LM Lexicon)	Constructed using corporate 10-K reports (financial articles) between 1994 and 2008.	Loughran and McDonald (2011)
Financial Stability Sentiment Dictionary (FS Lexicon)	Constructed using financial stability reports of 35 central banks between 2005 and 2015.	Correa et al. (2017)
Hu and Liu Opinion Lexicon (HL Lexicon)	Constructed by determining the sentiment scores of e-commerce reviews using Part-of-Speech tagging and feature identification.	Hu and Liu (2004)
VADER	Constructed to determine the intensity of social media text snippets (tweets) with annotated features.	Hutto and Gilbert (2014)
Genetic Algorithm-based Lexicon (PH Lexicon)	Constructed by using an algorithm that finds the combination of words that maximizes the number of correctly classified polarities from annotated text while minimizing the number of words.	Holland (1975) and Castelli et al. (2017)

Machine learning-based method

The ML-based approach predicts the sentiment of text using a pre-trained model. It is shown to capture more nuances in text but is more complex, requires more computational power, and is less transparent. We used the following ML models:

1. Bidirectional Encoder Representations from Transformers (BERT)

BERT (Devlin, 2018) is a transformer-based machine learning technique developed by Google for NLP tasks. BERT can capture the context and meaning of words based on both their preceding and succeeding words. On various sentiment analysis benchmarks and competitions, BERT has achieved state-of-the-art performance. Applications of BERT include sentiment analysis, social media monitoring, and others.

2. Financial Bidirectional Encoder Representations from Transformers (FinBERT)

FinBERT (Huang, A. H., Wang, H., and Yang, Y., 2022) is short for FinancialBERT, a domain-specific variant of BERT designated to understand and analyze financial data. It is specifically trained and fine-tuned on financial texts such as business articles from Reuters and corporate 10-K reports.

Sentiment Scoring and Model Selection

Between the dictionary-based and ML-based approaches in quantifying sentiment in text, we adopted a more straightforward method of scoring for the dictionary-based method, only using integer representations of positive, negative, and neutral sentiment (1, -1, and 0, respectively). The net diffusion of words determines the news sentiment for an article:

$$Sentiment\ Score_{article,dictionary} = \frac{\sum Positive_{w,article} - \sum Negative_{w,article}}{Total_{w,article}}$$

where *w, article* corresponds to the count of words in an article. Following Hutto & Gilbert (2014) and Asmi & Ishaya (2012), we have implemented a linguistic rule known as negation rule. This is where a word switches polarity (-1 to 1 or vice versa) whenever it is preceded by a negation word, such as 'not' or 'never'.

Given that the ML models return a probability score of sentiment, we assign the results to either positive, negative, or neutral. Since the models compute on a per-sentence basis, the news sentiment of an article is derived by the net diffusion of sentences:

$$Sentiment\ Score_{article,ML} = \frac{\sum Positive_{s,article} - \sum Negative_{s,article}}{Total_{s,article}}$$

where *s, article* corresponds to the count of sentences in an article.

Since the number of articles per news source varies per month, we followed the approaches of Chong et al., (2022) and Baker et al., (2016), where the NSIs are calculated from the average sentiment scores of the articles in each month, adjusted by the number of articles per news source. The indices were also standardized to a historical average of zero. Estimates of the index above zero indicate positive sentiment, and estimates below zero indicate otherwise.

A set of 3000 labeled sentences were used as a test dataset. The selected sentences were also taken from the same news data sources and contained at least one word from any of the five lexicons. The same dataset was used in constructing a Philippine-specific lexicon using a genetic algorithm (PH lexicon) that retrieves a combination of keywords that maximizes the number of correctly classified sentences from the test dataset. The words that comprise the PH lexicon can be found in Appendix A.

To select which lexicon to use for the main NSIs for the dictionary-based and machine learning-based approaches (NSI-D and NSI-ML, respectively), we assessed the classification performance of the five off-the-shelf lexicons along with the constructed PH lexicon and the two pre-trained machine learning models vis-à-vis the manually labeled dataset based on accuracy and macro-F1 scores.

Accuracy measures the fraction of the models' correct sentiment predictions. Meanwhile, macro-F1 score combines precision (a measure of false positives) and recall (a measure of false negatives). Macro-F1 is a commonly used metric in NLP literature (Shapiro et. al, 2020) and a function of F1 score with the number of classes taken into consideration:

$F1\ score = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2TP}{2TP+TP+FN}$ and $Macro\ F1\ score = \frac{1}{n}\sum_i^n F1\ score_i$

where *n* is the number of classes.

While accuracy is a commonly used metric in evaluating the performance of a model, it does not consider the distribution of the data and is adversely affected by false negatives. The macro-F1 score, on the other hand, penalizes models that have too many false positives and false negatives. In addition, macro-F1 score performs better with imbalanced classes. Thus, the macro-F1 score is a more appropriate measure for model evaluation.

The model with the highest accuracy and macro-F1 score is used to generate the NSI-D and NSI-ML. Table 3 shows the accuracy and macro-F1 scores of the various lexicons and ML models. As may be expected, the lexicon created using the genetic algorithm (PH lexicon) outperformed other lexicons, while FinBERT surpasses BERT in terms of accuracy and macro-F1 score. Thus, both the PH lexicon and FinBERT model were chosen to generate the NSI-D and NSI-ML, respectively.

Evaluation Metrics of Lexicons and Machine Learning Models

Table 3

Lexicons	Accuracy	Macro-F1 score
PH Lexicon	0.66	0.64
LM Lexicon	0.52	0.46
FS Lexicon	0.51	0.51
VADER	0.51	0.42
HL Lexicon	0.50	0.46
GI Lexicon	0.45	0.42
ML models		
FinBERT	0.66	0.63
BERT	0.58	0.45

Source: Authors' estimates

Note: These are computed relative to the annotated dataset.

Topic Modeling

We now uncover the themes that best characterize the collection of news articles. Topic modeling involves using an unsupervised ML technique to cluster text based on similarities. We evaluated two of the most used methods for topic modeling: 1) Latent Dirichlet Allocation (LDA) and 2) Non-Negative Matrix Factorization (NMF).

Latent Dirichlet Allocation

LDA (Blei et al., 2003) is a generative model that represents documents as mixtures of topics, and each topic is made up of word probabilities. There are two well-known implementations of LDA: (a) Variational inference (VI) and (b) Markov Chain Monte Carlo-Gibb's sampling/Collapsed Gibb's sampling. Gibb's sampling yields precise results, while VI cannot guarantee approximation even after convergence (Blei et al., 2018). However, this comes with a significant tradeoff - Gibb's sampling is slower and more computationally expensive than VI. Due to its precision, we chose to use the Gibb's sampling implementation of LDA in evaluating different topic modeling algorithms.

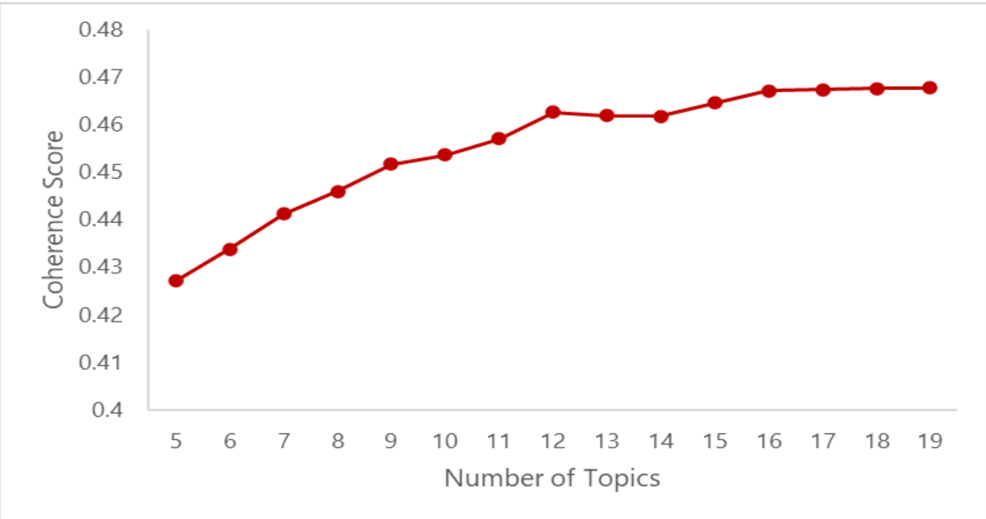
Non-Negative Matrix Factorization

Non-Negative Matrix Factorization (Lee and Seung, 2000) is a statistical method that decomposes a Term-Document matrix into Word-Topic and Topic-Document matrices. Given that NMF is an application of linear algebra, it is a swift and efficient topic modeling algorithm. The NMF model provided similar results to LDA while being less complex and taking less time to execute. Hence, we chose NMF over LDA to retrieve the topics from news articles.

Coherence Score

Coherence score is the degree of similarity between the top words of a certain topic. The graph of coherence score against the number of topics suggests a marginal increase for every additional topic. We computed the coherence score of topics ranging from 5 to 19, averaging the scores of 100 NMF models with different random states per number of topics. The coherence score is computed as the average of all the cosine similarities between each topic word and its cluster through the Python Gensim implementation derived from Roder et al. (2015). Figure 1 shows the coherence score of the overall coherence scores for NMF. Similar to a scree plot, we follow a heuristic where the first inflection point or the "elbow" in the curve is the optimal choice. In this case, we found that 12 is the optimal number of topics. Consequently, we still checked whether the resulting topics were consistent by looking at the keywords per topic.

Figure 1: Coherence scores against Number of Topics



Source: Authors' estimates

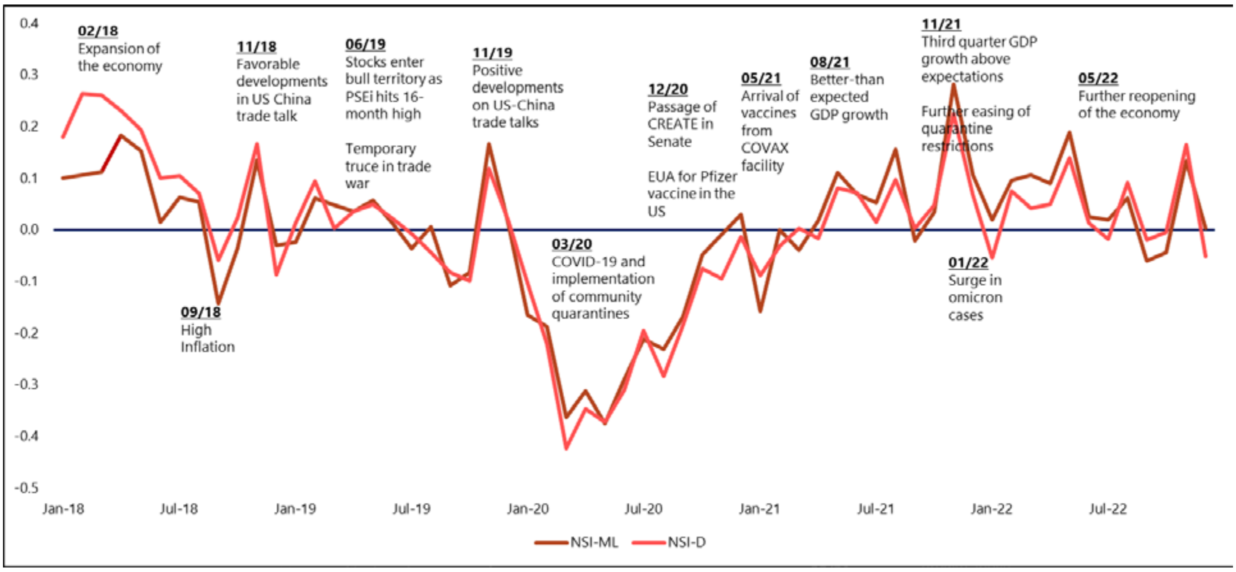
4. Results and Discussion

This section presents an analysis of the NSI-D and NSI-ML estimates. Figure 2 shows that the indices move together, suggesting broadly similar estimates even if these NSIs are derived using two different methodologies. More importantly, Figure 2 suggests that the NSI estimates match the expected market sentiment of key events that may have significant economic repercussions.

For example, sentiment dipped and turned significantly negative following the onset of the COVID-19 pandemic and the implementation of community quarantines. It marginally recovered in December 2020 (the first time the index registered a positive reading since the pandemic), coinciding with the passage of the Corporate Recovery and Tax Incentives for Enterprises (CREATE) bill in the Senate and the issuance of the emergency use authorization of the Pfizer vaccine in the US. Sentiment estimates were also above zero in May 2021 when the vaccines from the World Health Organization-led COVAX facility arrived in the Philippines. Sentiment estimates also settled in the positive region when favorable economic outturns, such as news of actual gross domestic product (GDP) exceeding expectations, were reported. Other key events that match the peaks (positive sentiment) and troughs (negative sentiment) of the NSIs have been identified in Figure 2 for the readers' reference. Overall, the results show that both NSI construction methods (NSI-D and NSI-ML) reflect economic upturns and downturns.

Figure 2: NSIs and Key Economic Events

January 2018 – December 2022



Source: Authors' estimates

Correlation of NSIs with Selected Economic Indicators

Table 4 presents the correlation of NSI-D and NSI-ML against selected economic indicators such as Philippine Stock Exchange Index (PSEi), Purchasing Managers'

Index (PMI) and the confidence indices (CI) from the BSP BES and CES. As expected, the NSI-D and NSI-ML are positively correlated with the PSEi, PMI, BES CI, and CES CI.

Correlation Scores of NSIs against Selected Economic Indicators

January 2018 – December 2022 Table 4

	PSEi	PMI	BES	CES
NSI-D	0.61* (0.00)	0.71* (0.00)	0.49** (0.03)	0.38 (0.10)
NSI-ML	0.53* (0.00)	0.67* (0.00)	0.43*** (0.07)	0.28 (0.23)

Source: Authors' Estimates

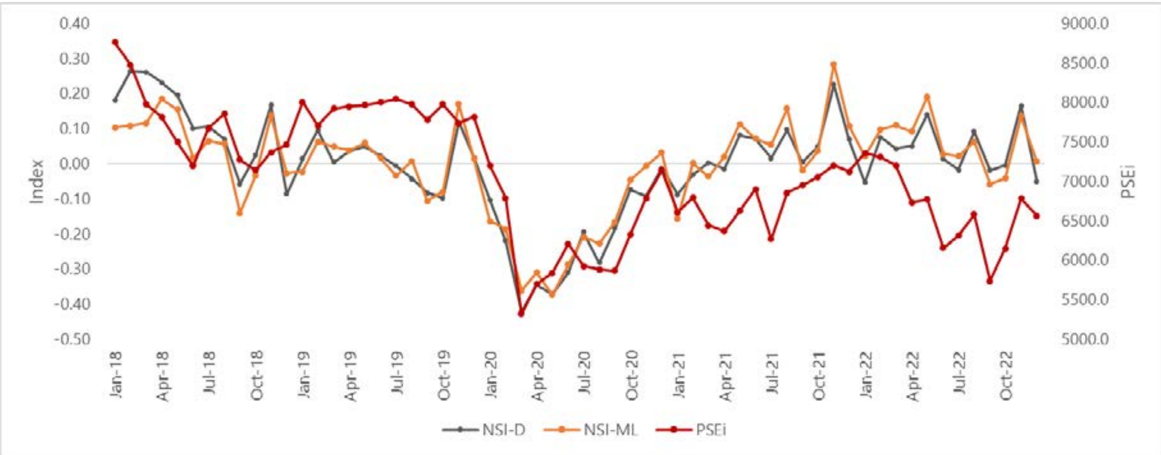
Note: Values in parenthesis are corresponding p-values. Asterisks *, **, and *** denote significance at 1%, 5% and 10% levels, respectively.

The monthly averages of NSI and selected economic indicators, such as the monthly average of daily PSEi and the monthly PMI, are shown in Figures 3 and 4, respectively. Both NSIs correlate positively with the PSEi at 0.61 and 0.53 for the NSI-D and NSI-ML, respectively.

The NSIs are also positively correlated with the PMI, which tracks economic or market conditions as seen by purchasing managers. Similar to the PSEi, NSI-D is more correlated with PMI at 0.71 compared to 0.67 of NSI-ML.

Figure 3: Correlation of NSIs with Average Monthly PSEi

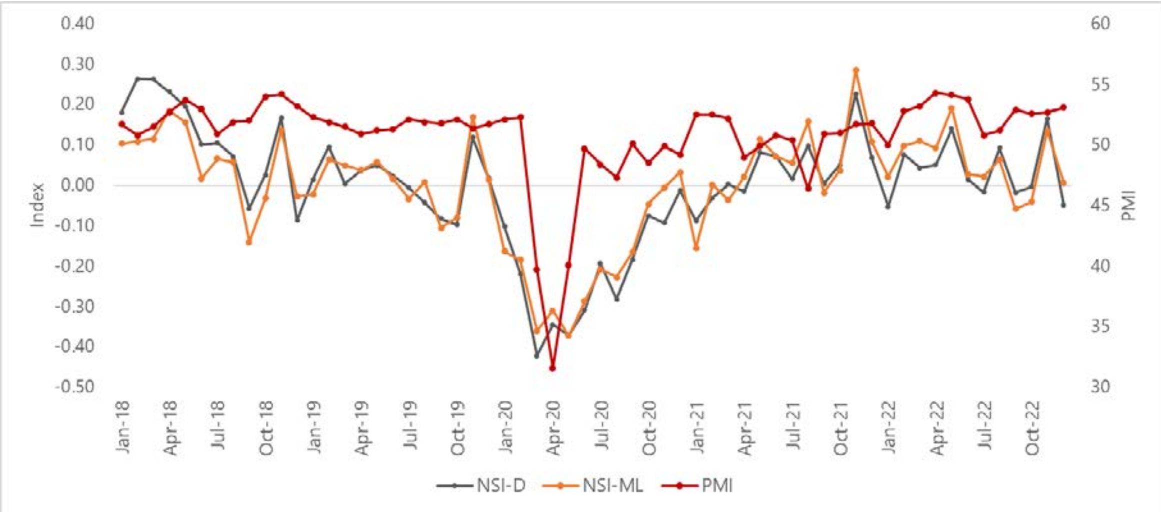
January 2018 – December 2022



Source: Authors' estimates

Figure 4: Correlation of NSIs with Monthly PMI

January 2018 – December 2022

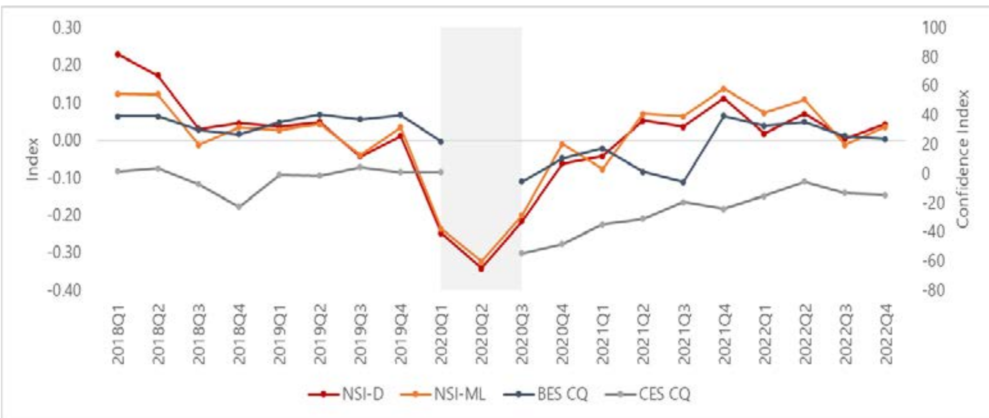


Source: Authors' estimates

To compare the NSIs with the quarterly economic indicators, such as the BES and CES current quarter confidence indices, we compute the quarterly average NSIs from Q1 2018 to Q4 2022, as shown in Figure 5. The quarterly NSIs show a positive correlation with business and consumer confidence indices. For the BES, NSI-D is higher at 0.49 compared to 0.43 with NSI-ML, while for the CES, the correlation with NSI-D is weaker at 0.38 and 0.29 with NSI-ML. Given that the NSIs are available at a higher frequency, this can fill the gaps when surveys are unavailable, like in Q2 of 2020 when survey activities were halted because of the COVID-19 pandemic.

Figure 5: Correlation of NSIs with Confidence Indices

2018Q1 -2022Q4



Source: Authors' estimates

Topic Modeling Results

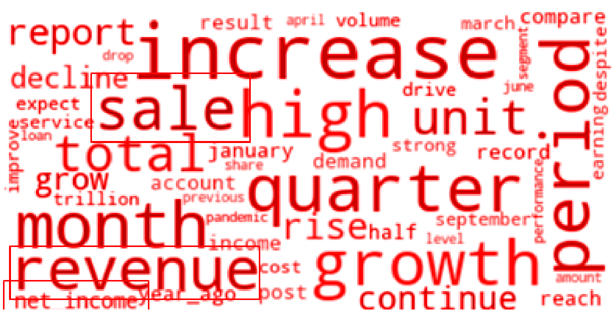
Based on the top words per topic generated through NMF, the following themes were identified: energy, agriculture, services, corporate news, stock market, micro, small and medium enterprises (MSMEs), banking, trade, government finance, government projects, company earnings, and inflation.

Figure 6 shows some of the word clouds associated with selected topics. Word clouds for all topics can be found in Appendix B. The word clouds and their respective notable keywords are as follows (clockwise starting from upper left): The first word cloud refers to Banking with the words "banking", "lending", "financial institution" alongside words pertaining to the BSP (e.g., "Bangko Sentral", "central bank", and "BSP"). For the next word cloud, emphasis is given to the words "inflation", "Federal Reserve" and "central bank" pertaining to monetary policy. Energy can be assigned as the topic in the next word cloud with the words "electricity", "power", "capacity", and "meralco" (Manila Electric Company) as keywords. Aside from the words "revenue", "sale" and "net income", frequency terms such as "month", "quarter" and specific months of year are usually observed with reports of publicly listed firms. Hence, we label the topic as Company Earnings.

Figure 6: Word clouds of selected topics

Banking

Monetary Policy



Company Earnings

Energy

Correlation of Topic-level NSIs against Economic Indicators

We analyzed how the NSI per topic or theme moves with economic indicators. Due to its relative ease of implementation and higher correlation scores with economic indicators compared to NSI-ML (see Table 2), NSI-D was used to generate separate NSIs for each of the identified topic. Correlation scores were also computed for NSI-D per topic with PSEi, PMI, BES, and CES CI. Table 5 shows the topics with which sentiment scores posted the highest correlation with the monthly PSEi and PMI.

Correlation Scores of Topic-level NSIs against Selected Economic Indicators

January 2018 – December 2022

Table 5

	PSEi	PMI
Company Earnings	0.70	0.54
Services	0.39	0.48
Trade	0.39	0.62
Banking	0.45	0.50
Infrastructure	0.41	0.34

Source: Authors' estimates

Note: Sentiment indices per topic were estimated using NSI-D. The calculated p-values are close to 0 (p < 0.01), hence significant at 1% level.

Company earnings (news articles relating to financial disclosures of firms) ranked highest among the topics with correlation against PSEi, followed by services and trade. Thus, it can be inferred that the NSI from news on company earnings can be used to gauge PSEi's movements and serve as a good indicator of market confidence. Meanwhile, market sentiment based on trade-related news (i.e., articles relating to imports and exports) appeared to be the most correlated with PMI, followed by company earnings and banking. The trends of the topic-level NSIs with the highest correlation against monthly PSEi and PMI are shown in Appendix C.

Table 6 shows the topics whose NSIs had the highest correlation scores with BES and CES CI. The NSI of news relating to company earnings ranked highest, followed by infrastructure and trade. This finding implies that market sentiment related to company earnings can also be an alternative measure of consumer and business outlook. Interestingly, these topics also displayed a high positive correlation with the GDP growth rate. The trends of the topic-level NSIs with the highest correlation against quarterly BES, CES, and GDP growth are shown in Appendix D.

Top Topics with NSIs that Exhibited Highest Correlation Scores with quarterly BES CI, CES CI and GDP growth rate

January 2018 – December 2022 Table 6

	BES CI	CES CI	GDP
Company Earnings	0.77	0.79	0.81
Infrastructure	0.60	0.62	0.70
Trade	0.47	0.36	0.78

Source: Authors' estimates

Note: Sentiment indices per topic were estimated using NSI-D. The calculated p-values are close to 0 (p < 0.01), hence significant at 1% level

It can also be noted that the above topic-level NSIs display a higher correlation with selected economic indicators compared to the general NSIs. This suggests that topic modeling may bring out specific facets of news that are more significant in explaining co-movement with selected economic indicators, particularly the BES, CES, and GDP growth. Therefore, these topic NSIs show potential as complements to these key economic measures.

8. Conclusion and Recommendations

News-based indices offer a faster and more cost-effective approach to high-frequency sentiment measurement by leveraging widely available news articles and innovations in sentiment analysis techniques.

Evaluation of the indices shows that the constructed NSIs match the expected sentiment for critical economic events. This suggests that NSIs have the potential to complement the BSP's existing BES and CES and provide a timely measurement of market sentiment.

Meanwhile, topic modeling helped identify the key themes in the corpus of news data used in this study. Notably, some topic-level NSIs demonstrated higher correlations with key economic indicators than the broad NSIs. This implies that topic modeling can help uncover specific news themes that may have more predictive content for key economic variables, particularly the BES, CES, and GDP growth rate. While correlation does not imply causality, the high correlation scores between the identified topics and key economic indicators are encouraging and suggest that this strand of research is worth pursuing.

If data limitations are addressed, further studies may generate a more extended series of NSIs to determine how it track economic indicators in the long run. To make our models more robust, we plan to fine-tune our existing models and explore new ones to improve the accuracy of classifying sentiments from texts. For example, more linguistic rules can be added (on top of the negation rule), and Part-of-Speech tagging can be considered for NSI-D. We also intend to improve NSI-ML by retraining FinBERT with the manually annotated dataset and complementing our analysis with other ML models (e.g., PH Lexicon + Support Vector Machines (SVM), FinBERT +

Random Forests). Other topic modeling techniques, such as dynamic topic modeling, may also be explored to account for the changes in topics over time.

References

Asmi, A., & Ishaya, T. (2012, October). Negation identification and calculation in sentiment analysis. In The second international conference on advances in information mining and management (pp. 1-7).

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. The quarterly journal of economics, 131(4), 1593-1636.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. Journal of the American statistical Association, 112(518), 859-877.

Chong, E., Ho, C. C., Ong, Z. F., & Ong, H.H, (2022). Using news sentiment for economic forecasting: a Malaysian case study. IFC Bulletins chapters, in: Bank for International Settlements (ed.), Machine learning in central banking, volume 57, Bank for International Settlements.

Correa, R., Garud, K., Londono, J. M., & Mislang, N. (2021). Sentiment in central banks' financial stability reports. Review of Finance, 25(1), 85-120.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805. Retrieved from https://arxiv.org/abs/1810.04805

Huang, Allen and Wang, Hui and Yang, Yi, FinBERT - A Large Language Model for Extracting Information from Financial Text (July 28, 2020). Contemporary Accounting Research, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3910214 or http://dx.doi.org/10.2139/ssrn.3910214

Holland, John (1992). Adaptation in Natural and Artificial Systems. Cambridge, MA: MIT Press. ISBN 978-0262581110.

Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In AAAI (Vol. 4, No. 4, pp. 755-760).

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Kanzari, D., Nakhli, M. S., Gaies, B., & Sahut, J. (2023). Predicting macro-financial instability – How relevant is sentiment? Evidence from long short-term memory networks. Research in International Business and Finance, 65, 101912. https://doi.org/10.1016/j.ribaf.2023.101912

Lee, D. D., & Seung, H. S. (2000). Algorithms for non-negative matrix factorization. In Advances in Neural Information Processing Systems (pp. 556-562).

Li, X., Ding, Y., Chen, X., & Lin, X. (2019). FinBERT: A Pretrained Language Model for Financial Communications. arXiv preprint arXiv:1908.10063. Retrieved from https://arxiv.org/abs/1908.10063

Marszal, Anna (2022). What news can really tell us? Evidence from a news-based sentiment index for financial markets analysis, NBP Working Papers 349, Narodowy Bank Polski.

Castelli, M., Manzoni, L., Vanneschi, L., & Popovič, A. (2017). An expert system for extracting knowledge from customers' reviews: The case of Amazon. com, Inc. Expert Systems with Applications, 84, 117-126.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of finance, 66(1), 35-65.

Nguyen, K., & Cava, G. L. (2020). Start Spreading the News: News Sentiment and Economic Activity in Australia. https://www.rba.gov.au

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. https://doi.org/10.1145/2684822.2685324

Seo, Beomseok (2022). Machine-Learning-Based News Sentiment Index (NSI) of Korea (한국의 기계학습 기반 뉴스심리지수) (2022). Bank of Korea WP 2022-15, Available at SSRN: https://ssrn.com/abstract=4230611 or http://dx.doi.org/10.2139/ssrn.4230611

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). Measuring News Sentiment. Federal Reserve Bank of San Francisco Working Paper 2017-01. Available at https://doi.org/10.24148/wp2017-01

Stevens, K., Kegelmeyer, P. W., Andrzejewski, D., & Buttler, D. (2012). Exploring Topic Coherence over Many Models and Many Topics. In Empirical Methods in Natural Language Processing (pp. 952–961). Available at: https://www.aclweb.org/anthology/D12-1087.pdf

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Appendix A

PH Lexicon Word List – First 300 Words

Polarity	Words
Positive[1]	able, absorb, absorbs, absorbing, absorbed, abuzz, acceleration, accustom, accustoms, accustoming, acquire, acquiring, acquired, actual, adherence, advance, advances, advancing, advanced, advantaging, advantaged, afford, affords, affording, afforded, agreeable, ahead expectations, ahead forecast, ally, allying, allied, amass, amasses, amassing, amassed, apex, appealable, appear, appears, appearing, assertion, assure, assures, assuring, attains, attaining, attractive, attractiveness, augment, augments, augmenting, augmented, auspicious, beat, beats, beating, beaten, beautiful, benign, better, betters, bettering, bettered, bloom, blooms, blooming, bloomed, bolster, bolsters, bolstering, bolstered, bonus, boost, boosts, boosting, boosted, brisk, build, builds, built, calm, calms, calming, calmed, canny, capability, careful, casual, ceaseless, champion, champions, championing, championed, charitable, cheer, cheers, cheering, cheered, clarification, clarity, classify, classifies, classifying, classified, clean, cleans, cleaned, collaborate, collaborates, collaborating, collaborated, collaborative, concentrate, concentrates, concentrating, concentrated, concession, conclusive, confident, conserve, conserves, conserving, conserved, constitution, contestability, convenience, conventional, convince, convinces, convincing, convinced, cool, cools, cooling, cooled, coordinate, coordinates, coordinating, coordinated, correction, cost reductions, courteous, creatively, culminate, culminates, culminating, culminated, dedicate, dedicates, dedicating, definite, deposit, depositing, deposited, despite, despites, despiting, despited, develop, develops, developed, difference, domain, dominance, driven strength, easier, ebullient, efficiencies, embrace, embraces, embraced, emerges, emerging, emphasize, emphasizes, emphasizing, emphasized, encouraged, encouragement, enjoys, enjoying, enjoyed, enlighten, enlightens, enlightening, enlightened, enter, entering, entered, entertain, entertains, entertaining, entertained, entices, enticing, enticed, equilibrium, essence, establish, establishes, establishing, established, esteem, esteems, esteeming, esteemed, evolve, evolves, evolving, evolved, excellent, excel, excels, excelling, excelled, exclusiveness, expansion, expect, expects, expecting, expectation, expenditure, expression, extent, faithful, fantastic, fastest, fastgrowing, fate, fates, fating, fated, favoring, firm, firming, flare, flares, flaring, flared, flexibility, freehold, freer, frontrunner, fuse, gain, gains, gaining, gained, generous, giddy, gift, gifts, gifting, gifted, give, gives, globalized, goal, grow, growing, grew, growth, guidance, harvests, harvesting, harvested, hearten, heartens, heartening, heartened, heroic, higher, hire, hires, hiring, hitch, hitches, hitching, hitched, illuminate, illuminates, illuminating, illuminated, imaginable, impenetrable, implore, implores, imploring, implored, improves, improved, include, increase, increased, inexhaustible, innovate, innovates, innovating, innovated, inquire, inquires…

[1] Displays only 300 out of 635 words of the lexicon

PH Lexicon Word List – First 300 Words (continued)

Polarity	Words
Negative[2]	abdicate, abdicates, abdicating, abdicated, aberrant, abet, abets, abetting, abetted, abnormal, abnormally, abolish, abolishes, abolished, abort, aborts, aborting, aborted, abrogate, abrogates, abrogating, abrogated, abrogations, abrupt, absences, abusive, abusively, accidental, accomplice, accuse, acquiesce, acquiesces, acquiescing, acquiesced, addict, addicts, addicting, addicted, addiction, adulteration, adversities, aftermaths, agonize, agonizes, agonizing, agonized, ail, ails, ailed, alarm, alarms, alarming, alarmed, alien, aliening, aliened, allegation, allegations, allege, alleges, amortization, angry, anguish, anguishes, anguishing, anguished, annoy, annoys, annoying, annoyed, annulment, annulments, annul, annuls, annulling, annulled, anomalously, anticompetitive, anxious, appall, appals, appalling, appalled, apprehensive, arbitrarily, arbitrariness, arbitrary, argue, argues, arrear, arrears, arrests, arresting, arrogance, arrogant, assault, assaults, assaulting, assaulted, attrition, avaricious, backbreaking, backdate, backdates, backdating, backdated, backward, bailout, bald, balked, bandit, banish, banishes, banishing, banished, bankrupt, bankrupts, bankrupting, bankrupted, barbaric, batter, batters, battering, battered, below, bemoan, bemoans, bemoaning, bemoaned, bitterness, bizarre, blackspot, bland, blink, blinks, blinking, blinked, blood, bloods, blooding, blooded, bloodbath, blot, blots, blotting, blotted, blunder, blunders, blundering, blundered, blur, blurs, blurring, blurred, bog, bore, bores, boring, bored, bottom, bottoms, bottomed, boycott, boycotts, boycotting, boycotted, breach, breaching, breached, breakages, breakdowns, breaks, broke, breakup, brexit, bribe, bribes, bribing, bribed, brood, broods, brooding, brooded, browbeat, browbeats, browbeating, browbeaten, bum, bums, bumming, bummed, bumpy, burdening, bust, busts, busting, busted, cancellation, cancer, carelessness, catastrophes, catastrophic, cause, causes, caused, cautions, cautioning, cavernous, cease, ceases, ceasing, ceased, charge, charged, chasten, chastens, chastening, chastened, chide, chides, chiding, chided, chill, chills, chilling, chilled, choke, chokes, choking, choked, choosy, circumvent, circumvents, circumventing, circumvented, clash, clashes, clashing, clashed, clawback, climbdown, closure, closures, closuring, closured, clueless, clumsy, coercion, collisions, collude, colludes, colluding, colluded, combat, combats, combating, combated, complain, complains, complaining, complaints, complicate, complicates, complicating, complicated, concern, concerns, concerning, concerned, conciliate, conciliates, conciliating, conciliated, condemnations, condemns, condemning, condemned, confess, confesses, confessing, confessed, confession, confine, confining, confined, confinement, confinements, confiscate, confiscates, confiscating, confiscated, confiscation, confrontations, confronts, confronting, confuse, confuses, confused, consternation, constrict, constricts…

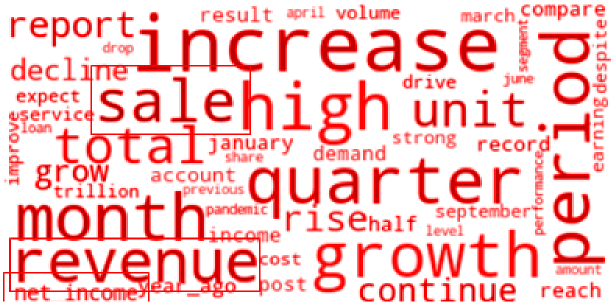
[2] Displays only 300 out of 2573 words of the lexicon

Appendix B

Word Cloud of Topics

Banking

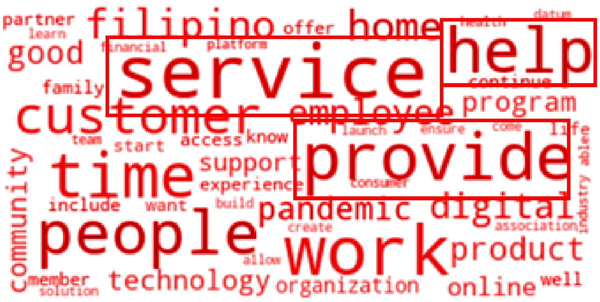
Monetary Policy



Company Earnings

Energy

Agriculture



Services

Word Cloud of Topics (continued)

Company News

Stock Market

MSMEs

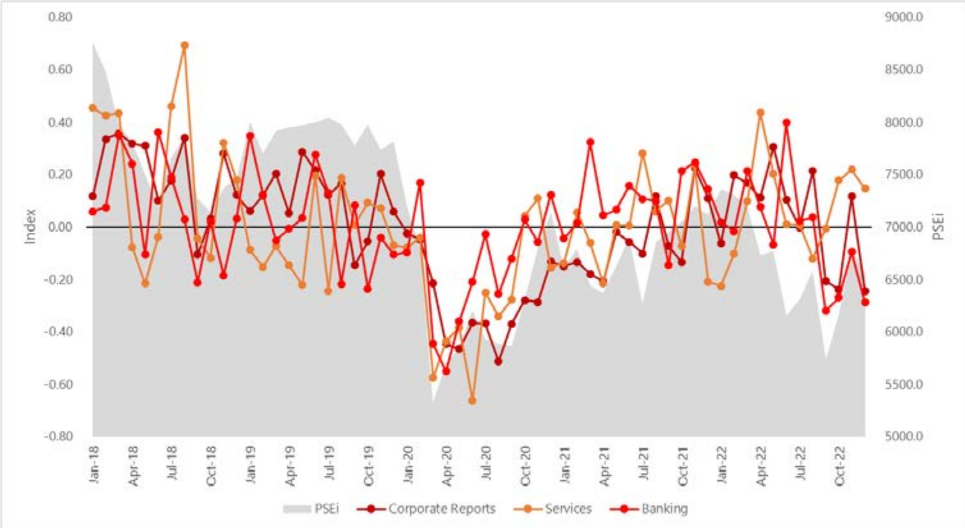
Trade

Fiscal Policy

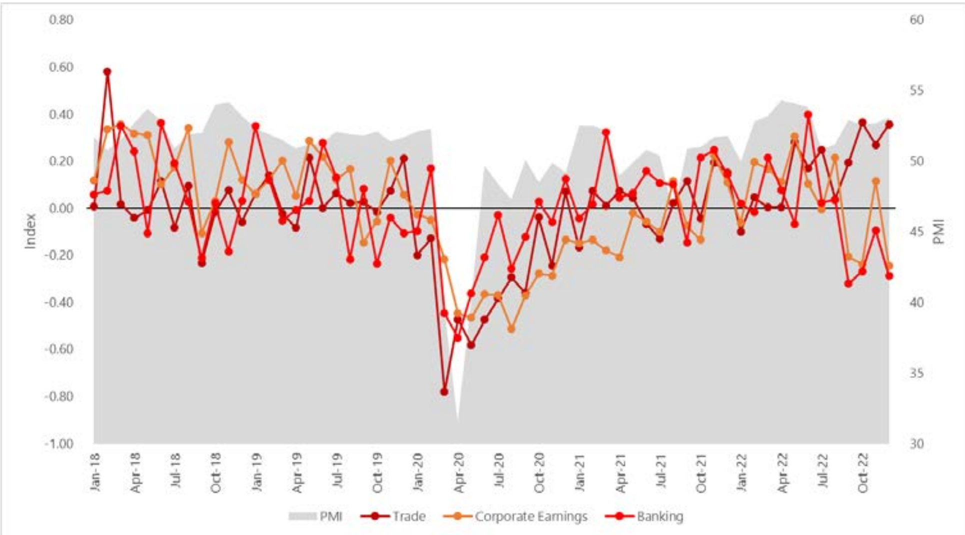
Infrastructure

Appendix C

Topic-level NSIs with Highest Correlation against Monthly PSEi

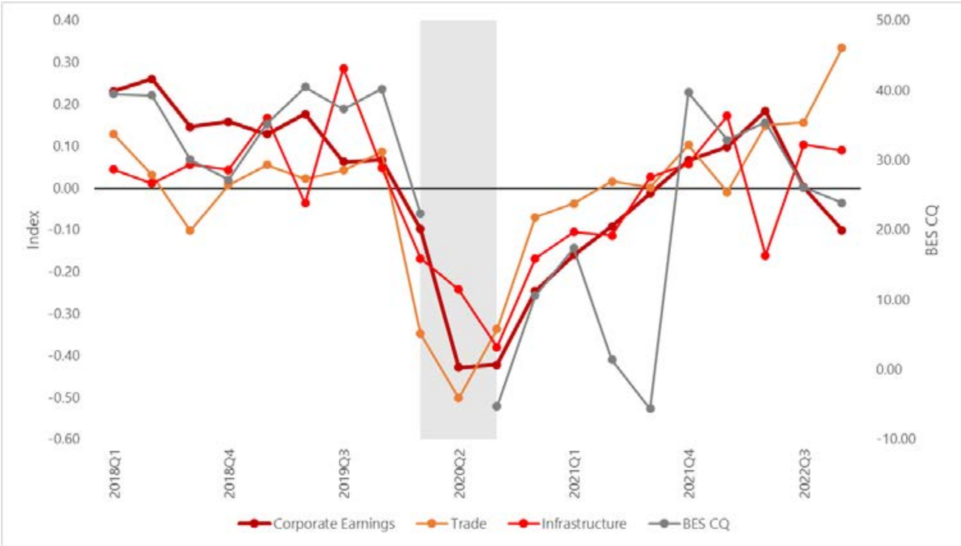


Topic-level NSIs with Highest Correlation against Monthly PMI

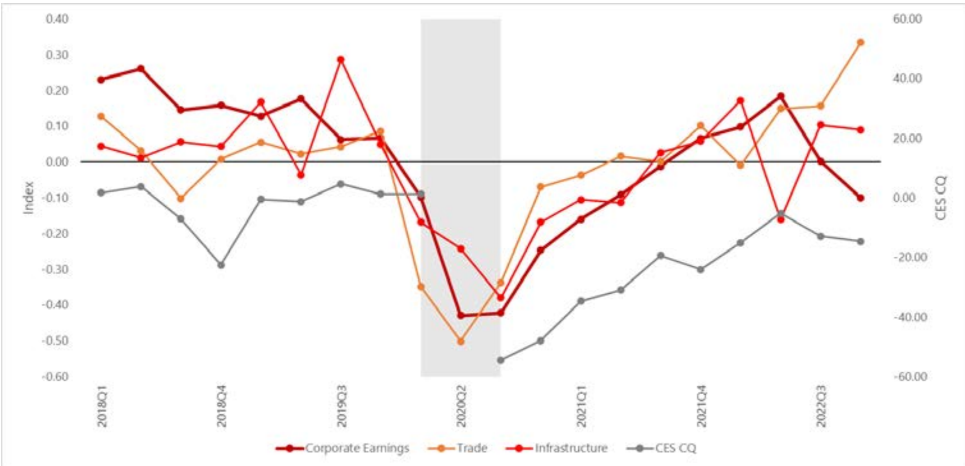


Appendix D

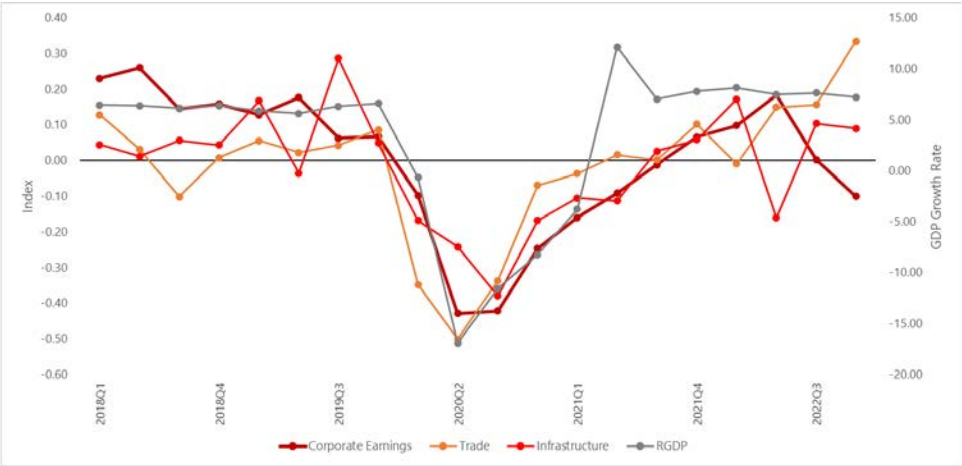
Topic-level NSIs with Highest Correlation against BES Current Quarter



Topic-level NSIs with Highest Correlation against CES Current Quarter



Topic-level NSIs with Highest Correlation against GDP Growth Rate



Constructing High-frequency and Thematic Economic Sentiment Indicators from News Articles

ALAN CHESTER ARCIN, CARMELITA ESCLANDA, CHELSEA ANNE ONG, ROSSVERN REYES

3RD IFC AND BANK OF ITALY WORKSHOP ON "DATA SCIENCE IN CENTRAL BANKING: ENHANCING THE ACCESS TO AND SHARING OF DATA", OCTOBER 17-19, 2023

Outline

1. Motivation
2. Overview, Data Collection and Pre-processing
3. Methodology
4. Model Evaluation
5. Results – Sentiment Indices
6. Topic Modelling
7. Results – Topic modelling
8. Key Takeaways and Future work

Motivation

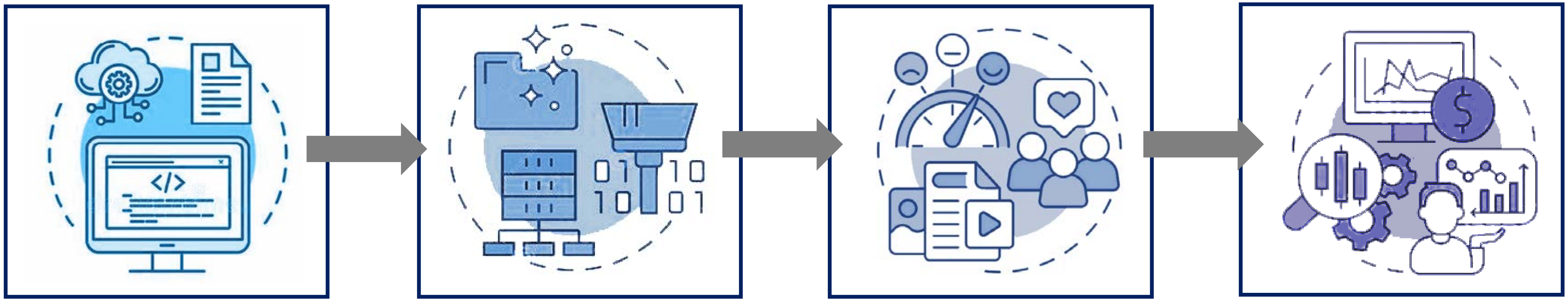
Sentiment influences economic agents' decisions

Survey-based measures of sentiment are **costly, tedious, and published with a delay**

News-based indices can be **produced at a higher frequency**, allowing for more up to date assessment at a lower cost

Overview

Web scraping → Data cleaning → Sentiment analysis → NSI



Data Collection and Pre-processing

Data were collected from multiple local news sources. The sections of interest were limited to economy, banking, finance and related fields. Text from the articles are pre-processed and further analyzed.

Pre-processing steps include:

- Removal of extra spaces in the text
- Case normalization
- Removal of punctuations and numbers
- Unicode conversion
- Tokenization
- Vectorization

Methodology

Dictionary based method involves a pre-defined list of keywords from some of the most used dictionaries in finance and well-known general dictionaries:

- General Inquirer (GI Lexicon)
- Loughran-McDonald Sentiment Dictionary (LM Lexicon)
- Financial Stability Sentiment Dictionary (FD Lexicon)
- Hu and Liu Opinion Lexicon (HL Lexicon)
- Valence Aware Dictionary for Sentiment Reasoning (VADER)

Using genetic algorithm, we retrieved combinations of keywords that maximizes the number of classified sentences from ~3000 manually labelled sentences (PH Lexicon).

Methodology

Machine learning based method utilizes two pretrained models to predict sentiment of text:

- BERT
  - ML model specifically trained using Wikipedia and BooksCorpus for common language tasks

- FinBERT
  - Built by further training BERT using financial text such as 10-K and 10-Q Corporate Reports, analyst reports and fine tuning it for financial sentiment classification

Model Evaluation

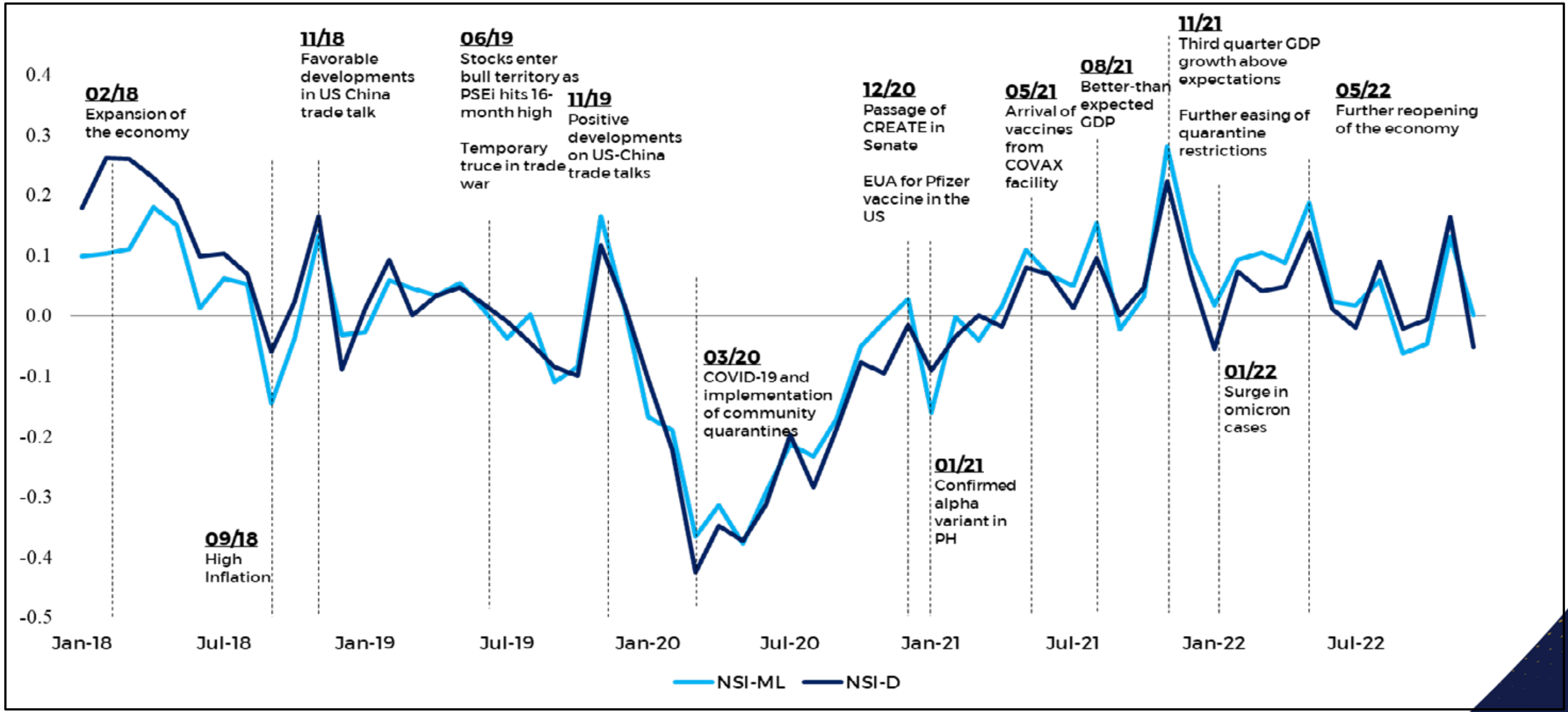
Both dictionaries and machine learning models were evaluated against ~3000 manually labelled sentences sampled from online news articles

Lexicons	Accuracy	Macro F1
PH Lexicon	0.66	0.64
LM Lexicon	0.52	0.46
FD Lexicon	0.51	0.51
VADER	0.51	0.42
HL Lexicon	0.50	0.46
GI Lexicon	0.45	0.42
Machine Learning Models	**Accuracy**	**Macro F1**
FinBERT	0.66	0.63
BERT	0.58	0.45

Accuracy measures the fraction of the model's correct predictions over total predictions. Meanwhile, macro F1 score combines both precision (a measure of the proportion of true positives that were correct) and recall (a measure of the proportion of true positives that were actually predicted).

Results

Estimates of NSIs strongly co-move with key economic events.



Results

Correlation analysis suggest co-movements between NSIs and selected economic indicators

Model	PSEi	PMI	BES	CES
NSI-D	0.61	0.62	0.49	0.38
NSI-ML	0.53	0.58	0.43	0.28

The Business Expectations Survey (BES) is one of the quarterly surveys conducted by the Bangko Sentral ng Pilipinas (BSP) which gathers information about entrepreneurs' views on the general business situation in their own industry and on the national economy. Likewise, the Consumer Expectations Survey (CES) captures the economic outlook of consumers as an indication of the country's future economic condition.

Topic Modeling and Correlation Analysis

Topic Modeling involves unsupervised machine learning techniques to **cluster text based on similarities**.

We considered the two of the commonly used techniques in topic modelling:

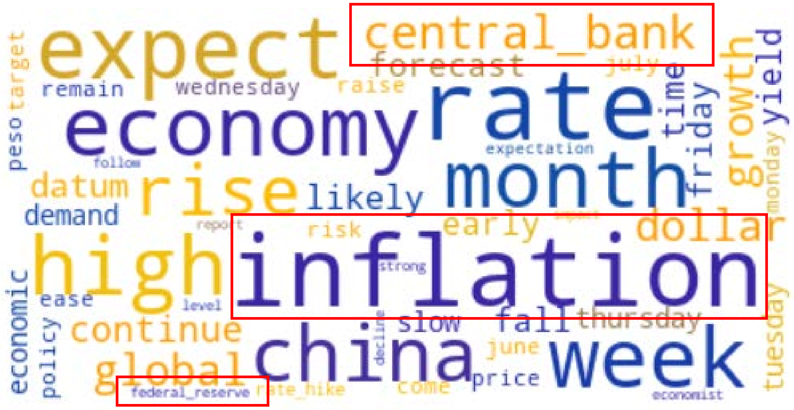
- **Non-Negative Matrix Factorization (NMF)**
- Latent Dirichlet Allocation (LDA)

Using the coherence score, we determined **12** to be the optimal number of topics. We then group the articles according to topics, compute the NSI for each and correlate these with selected economic indicators.

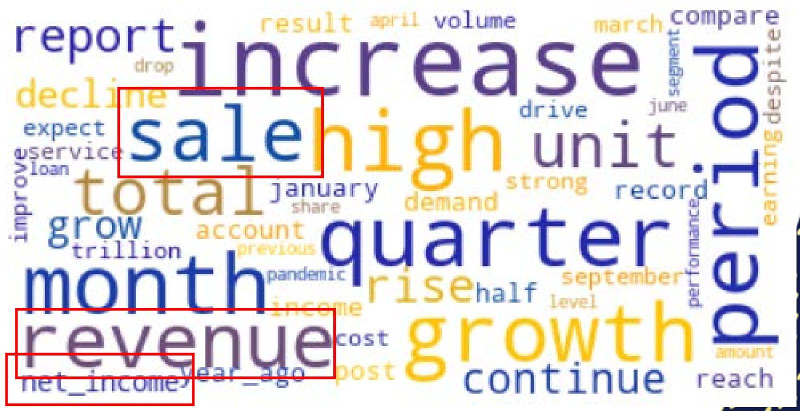
Results

Word clouds of selected topics

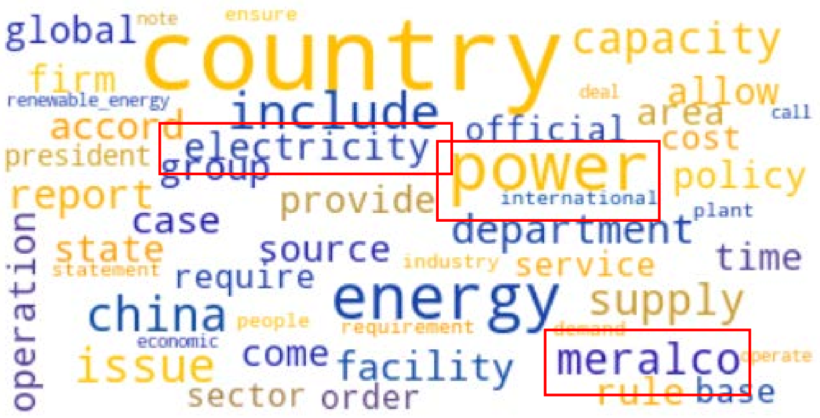
Banking



Monetary Policy



Company Earnings



Energy

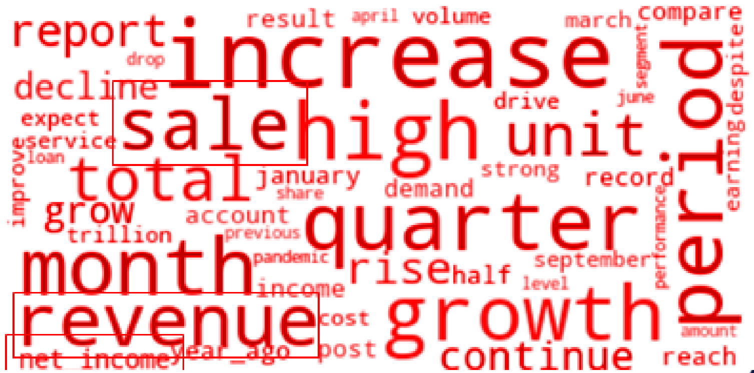
Agriculture

Results

Word clouds of selected topics

Banking

Monetary Policy

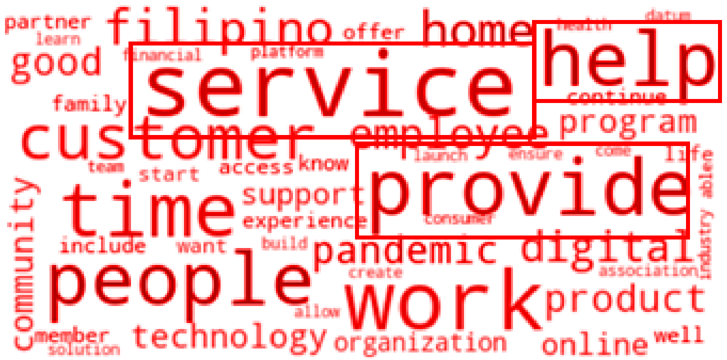


Company Earnings

Energy

Agriculture

Results



Services

Corporate News

Stock Market

MSMEs

Philippine Trade

Fiscal Policy

Infrastructure

Results

Correlation analysis show comovements of selected topics against monthly indicators PSEi and PMI…

Topic	PSEi	PMI
Company Earnings	0.72	0.53
Services	0.45	0.47
Trade	0.44	0.63
Banking	0.41	0.49
Government Projects	0.44	0.35

As well as against quarterly indicators BES, CES and GDP growth rate

Topic	BES CI	CES CI	GDP
Company Earnings	0.74	0.79	0.81
Infrastructure	0.60	0.62	0.70
Trade	0.48	0.36	0.78

Key Findings and Future Works

Key Takeaways:

- NSIs are found to **match expected sentiment for key economic events**
- NSIs showed potential in **complementing existing surveys** (BES and CES) and providing **timely measurement of market sentiment**
- Topic modeling can **help uncover specific news themes** that may have more predictive content for specific key economic variables

Future Works:

- Enhance NSI-D (e.g., Add more linguistic rules, Part-of-Speech Tagging)
- Retrain FinBERT using the manually labelled dataset for NSI-ML
- Complement with other ML models (e.g., LM Lexicon + SVM, FinBERT + Random Forests, etc.)

Sources

- Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis. M.I.T. Press.
- Loughran, Tim and McDonald, Bill, When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks (March 4, 2010). Journal of Finance, Forthcoming, Available at SSRN: https://ssrn.com/abstract=1331573
- Correa, Ricardo, Keshav Garud, Juan M. Londono, and Nathan Mislang (2017). Sentiment in Central Banks' Financial Stability Reports. International Finance Discussion Papers 1203.
- Zhang, L., Liu, B. (2017). Sentiment Analysis and Opinion Mining. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7687-1_907
- Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805. Retrieved from https://arxiv.org/abs/1810.04805
- Huang, Allen and Wang, Hui and Yang, Yi, FinBERT - A Large Language Model for Extracting Information from Financial Text (July 28, 2020). Contemporary Accounting Research, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3910214 or http://dx.doi.org/10.2139/ssrn.3910214

Thank you!

* The views expressed herein are those of the authors' only and do not necessarily reflect those of the Bangko Sentral ng Pilipinas