Hot off the press: An inflation news index for the Philippines using reinforced lexicons and memory-based deep learning

Ma. Ellysah Joy T. Guliman, Alan Chester T. Arcin, Jacqueline Margaux G. Herbo, Sanjeev A. Parmanand and Genna Paola S. Centeno, Cherrie R. Mapa

Bangko Sentral ng Pilipinas

CBSL International Research Conference
13 December 2024

**The views expressed in this paper are those of the authors and do not necessarily reflect those of the BSP. Any errors and omissions are the sole responsibility of the authors.*

Objectives and Motivation

This study is motivated by the following information:

- The news contain information that may influence the decision and behavior of economic agents
- Inflation expectations of consumers and investors could drive the future path of actual inflation
- Measures of inflation expectations have been mostly survey-based, which is costly and comes at a lag

Research Objectives:

1. To **construct** an additional high-frequency indicator of macroeconomic monitoring and surveillance
2. To **complement** the actual monthly consumer price index (CPI) and BSP Survey of External Forecasters (BSEF).

The Inflation News Index (INI) is a quantitative summary of digital news coverage on inflation and prices of goods and services.

Data Sources and Annotation

Data Sources

BusinessWorld

MANILA BULLETIN

INQUIRER.NET

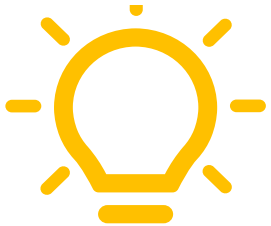
Manila Standard

BusinessMirror
A broader look at today's business

Annotation of sentences

Total: 3000 sentences
Class 1: Increase: 1,803
Class 2: Decrease: 659
Class 3: Unchanged/Irrelevant: 538

Problem: Imbalanced Classes



Solution:
Modify train and test datasets (Undersampling)

- Train dataset : 1000 (inc), 500 (dec), 500(no change/irrelevant)
- Test dataset: 803 (inc), 159 (dec), 38 (no change/irrelevant)

	Train dataset	Test dataset
Increase	1000	803
Decrease/No change	1000	197

SMOTE (Oversampling)
This technique synthesizes new samples of the minority class based on the nearest neighbors in the feature space.

Inflation News Indices

Inflation News Indices (INIs)

Lexicons

- Pointwise Mutual Information (PMI)
- Reinforcement Learning (RL)

INI-RL

- Bidirectional Long Short Term Memory (BiLSTM)

Machine Learning

INI-BiLSTM

Sentence Sentiment Labelling

1 Dictionary-based approach

- Pointwise Mutual Information

  - Measure of association between two words or between words and classes
  - $PMI(w|c) = \log\left(\frac{p(w|c)}{p(w)*p(c)}\right)$
  - Overall PMI score

$$-Score(w) = PMI(w|increase) - PMI(w|decrease)$$

Data Cleaning Processing

Pre-processing Procedures

1. Exclusion of ads and sponsored articles
2. Deletion of “lede” (the first paragraph summarizing the article)
3. Removal of extra spaces in the text
4. Conversion to lowercase
5. Removal of unwanted punctuation marks
6. Conversion of Unicode characters to ASCII
7. Tokenization of articles into sentences/words

Additional Procedures

Part-of-Speech (POS) Tagging

Shortage VBN

Surging JJ

Named Entity Recognition

Dakila PERSON

BSP ORG

Dictionaries

Initial Dictionary (Top 300 PMI Scores)

Words associated with decrease

slower, pulled, declines, appreciation, downward, tariffication, arrival, decline, benefit, downtrend, slowed, decreased, ….

Words associated with increase

depreciation, shortage, triggered, potential, reform, conflict, implement, premium, region, rising, faster, additional, …

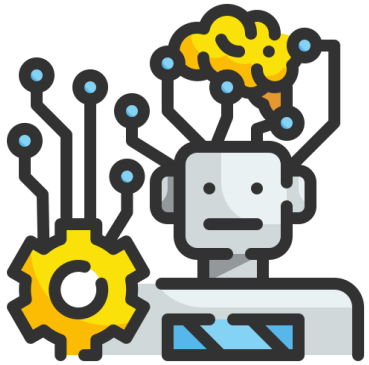
Dictionary w/ Reinforcement Learning

slower, pulled, declines, downward, tariffication, decline, downtrend, slowed, reduction, eased, decrease, softening, …

depreciation, shortage, triggered, potential, conflict, implement, premium, rising, faster, additional, upward, continuing, …

Sentence Sentiment Labelling

2 Machine Learning-based approach

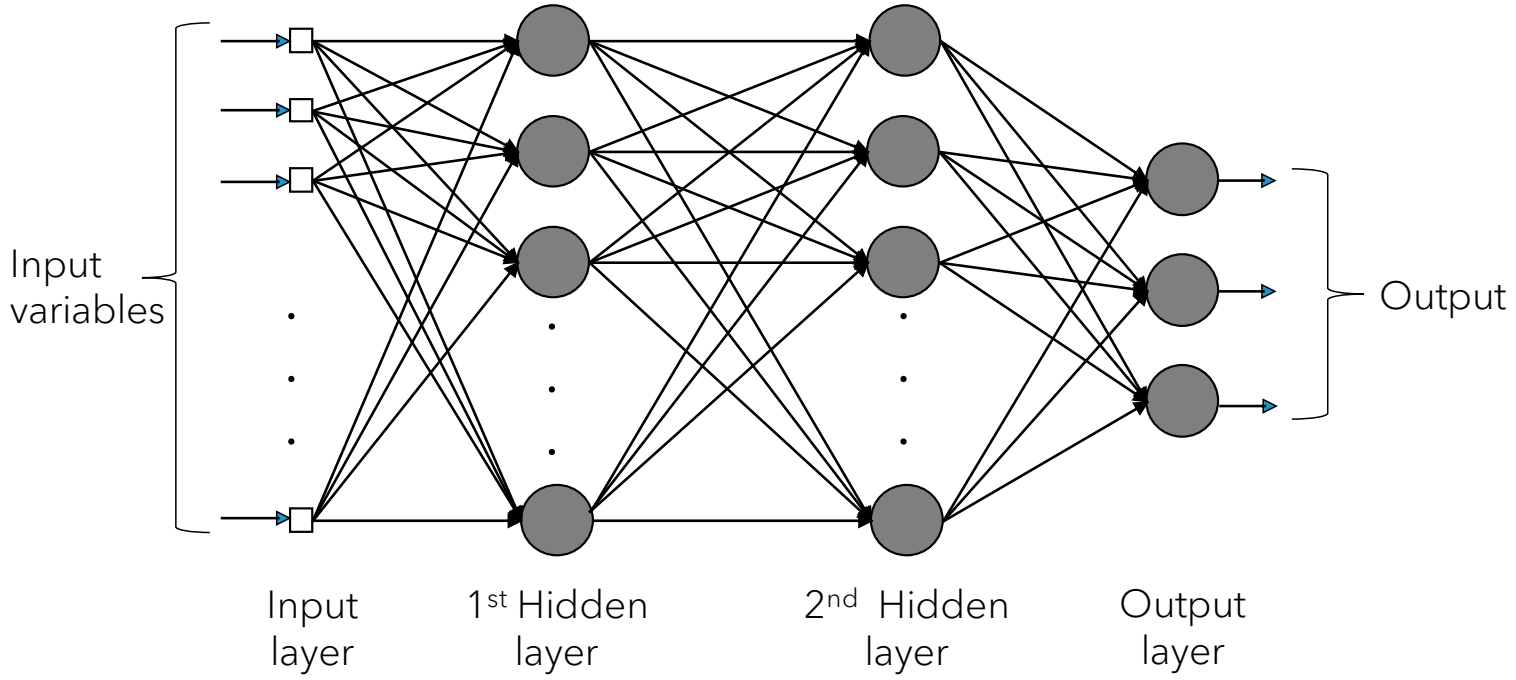


- Data Pre-processing
  - Removal of html tags and punctuations, lowercase, remove stopwords, lemmatization
- Word2Vec Embeddings
  - Vectorize the sentences by converting the words into numerical representations.
- Models explored: Artificial Neural Network (ANN), Bidirectional Long Short-Term Memory (Bi-LSTM)

Multilayer Perceptron (MLP)

Artificial Neural Network (ANN)

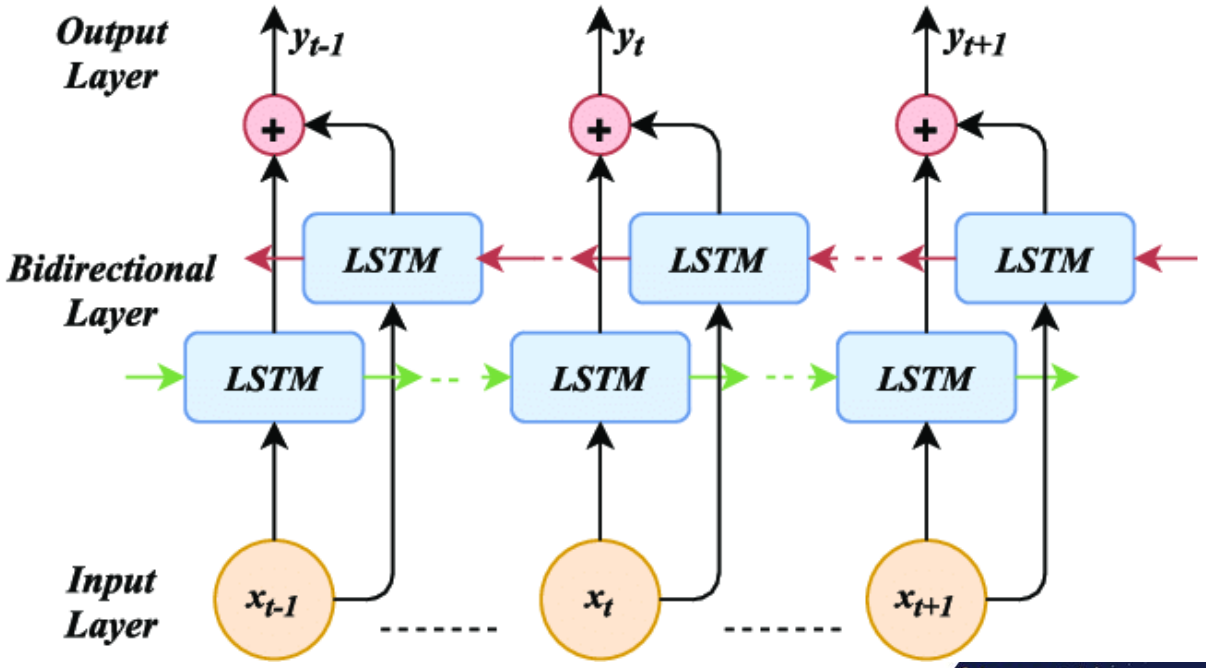
A three-layer perceptron



Source: Authors' own presentation

Bidirectional LSTM model

- Contains two LSTMs:
  - forward and backward direction
  - To learn relationship of preceding and following words



Evaluation of Methods

Lexicons and ML models were evaluated against the manually annotated dataset with around 3,000 sampled sentences from online news articles.

Dictionary*	Accuracy	Macro F1
Initial Lexicon	64.3	41.8
RL Lexicon	**89.3**	**69.3**

ML	Accuracy	Macro F1
MLP (20,8,3)	68.9	45.7
BiLSTM (16)	73.5	54.8
BiLSTM (32)	78.5	57.1
BiLSTM (16) SMOTE	78.4	56.2
BiLSTM(32) SMOTE	**79.8**	**57.9**

Source: Authors' calculations

NOTES: The MLP has input size of 20 (corresponding to Word2Vec embedding size), 8 hidden units in the first layer, and 3 neurons in the last layer (corresponding to the 3 classes: increase/decrease/no change).
A BiLSTM (x) notation is interpreted as a BiLSTM with x number of hidden units per LSTM. For example, BiLSTM(16) has 2 LSTM modules with 16 hidden units each.

Methodology: Index construction

- Index Construction Methods:

Method	Equation
1	$\mathrm{INI} = \frac{\#(articles\ net\ increase)}{\#(relevant\ articles)}$
2	$Score = \frac{\#(articles\ net\ increase) - \#(articles\ net\ decrease)}{\#(relevant\ articles)}$

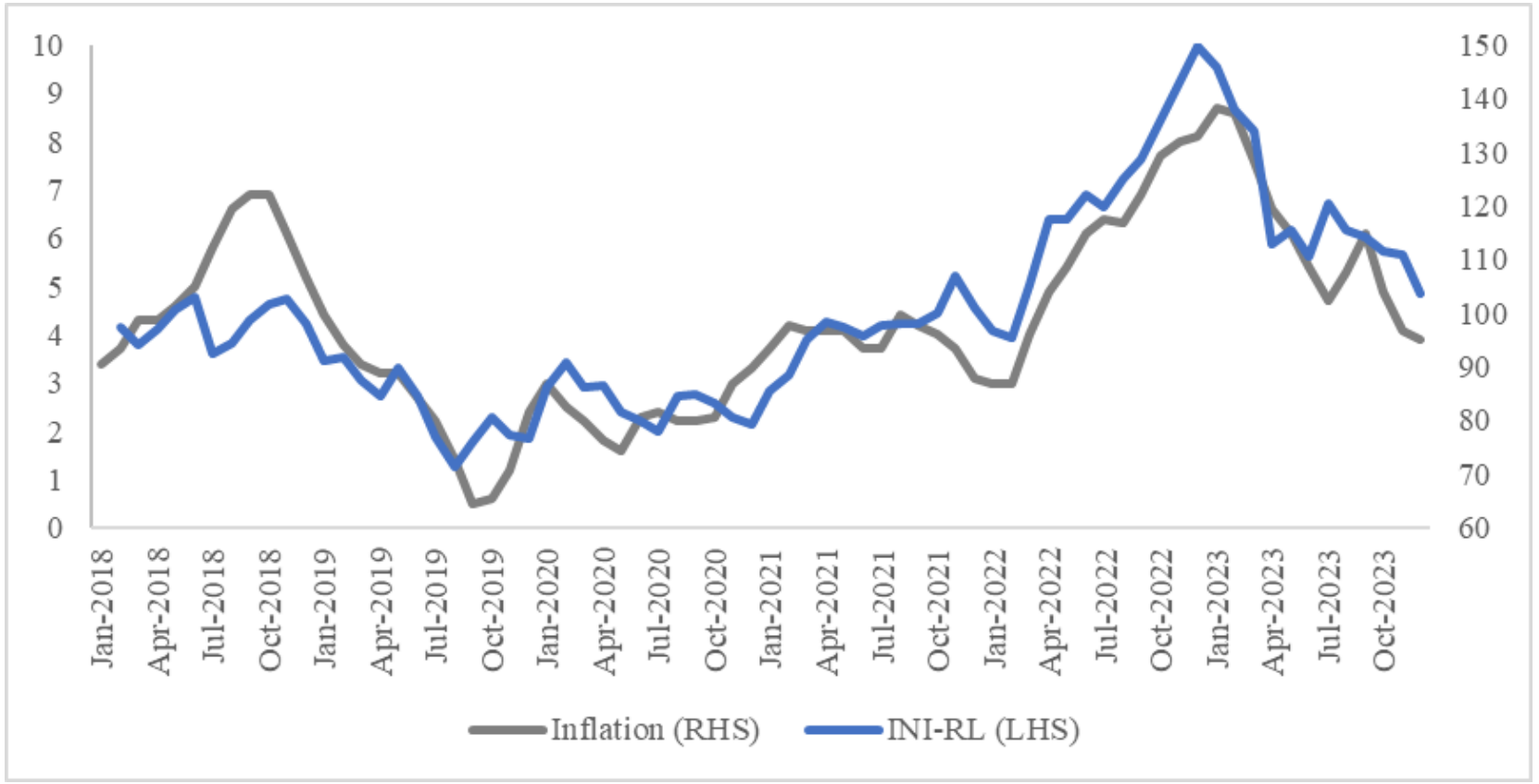
Evaluation of Results

Comparison of model accuracy and correlation with CPI

Model	Indexing Method	Test Accuracy (in percent)	Macro-F1 score	Correlation with CPI 2018	Correlation with CPI 2018 lag 1
BiLSTM (16)	1	73.5	54.8	0.731	0.756
BiLSTM(32)	1	78.5	57.1	0.839	0.849
BiLSTM (16) with SMOTE	1	78.4	56.2	0.813	0.827
BiLSTM (32) with SMOTE	1	79.8	57.9	0.820	0.838
RL lexicon	2	89.3	69.3	0.822	0.844

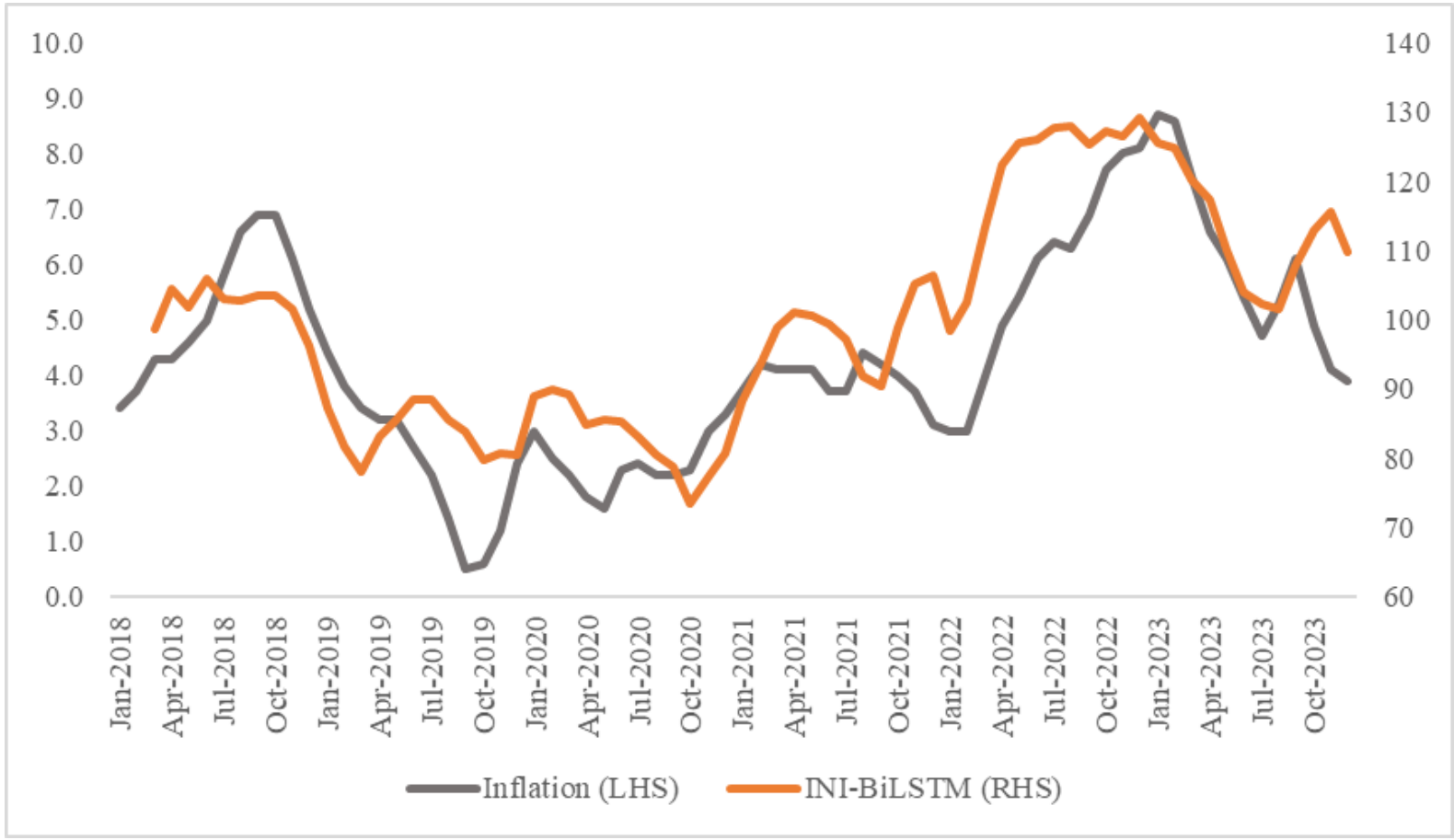
Preliminary Results

INI using the RL lexicon vs. CPI base 2018

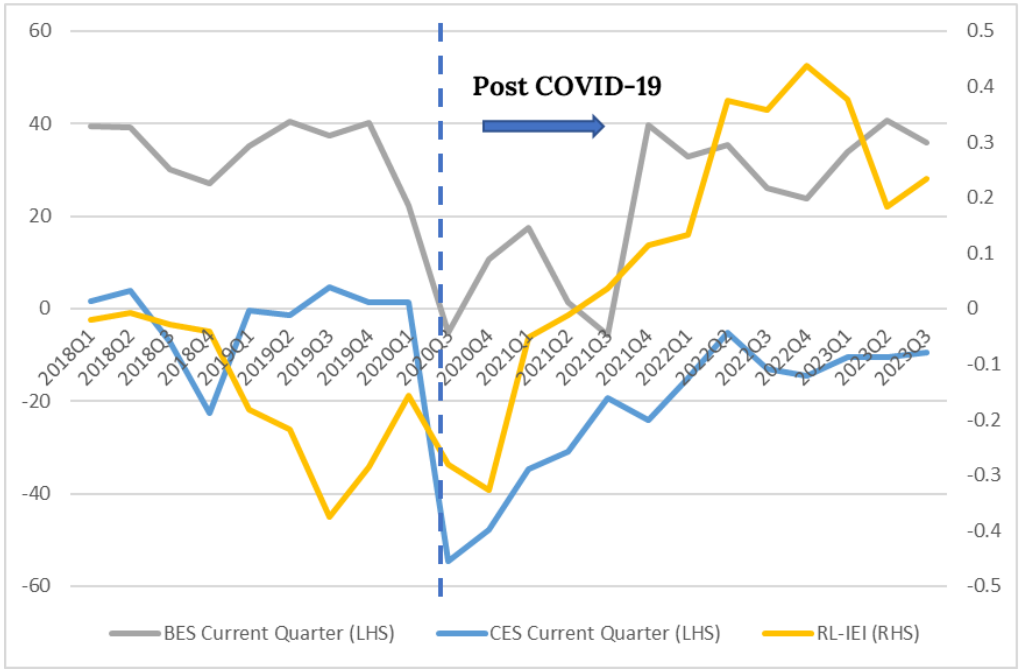


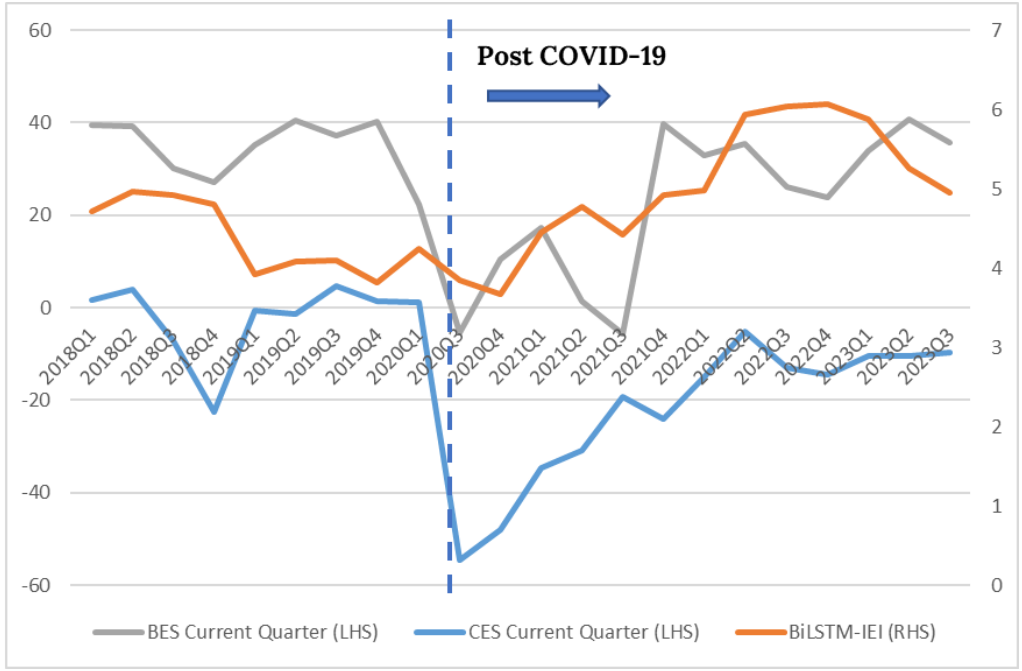
Preliminary Results

INI using Bidirectional LSTM vs. CPI base 2018



INIs vs Inflation Expectations





Correlation between INIs and Business/Consumer Expectations

	BES-CQ*	BES-NQ*	CES-CQ**	CES-NQ**
INI-RL	0.82	0.83	0.57	0.51
INI-BiLSTM	0.70	0.71	0.30	0.23

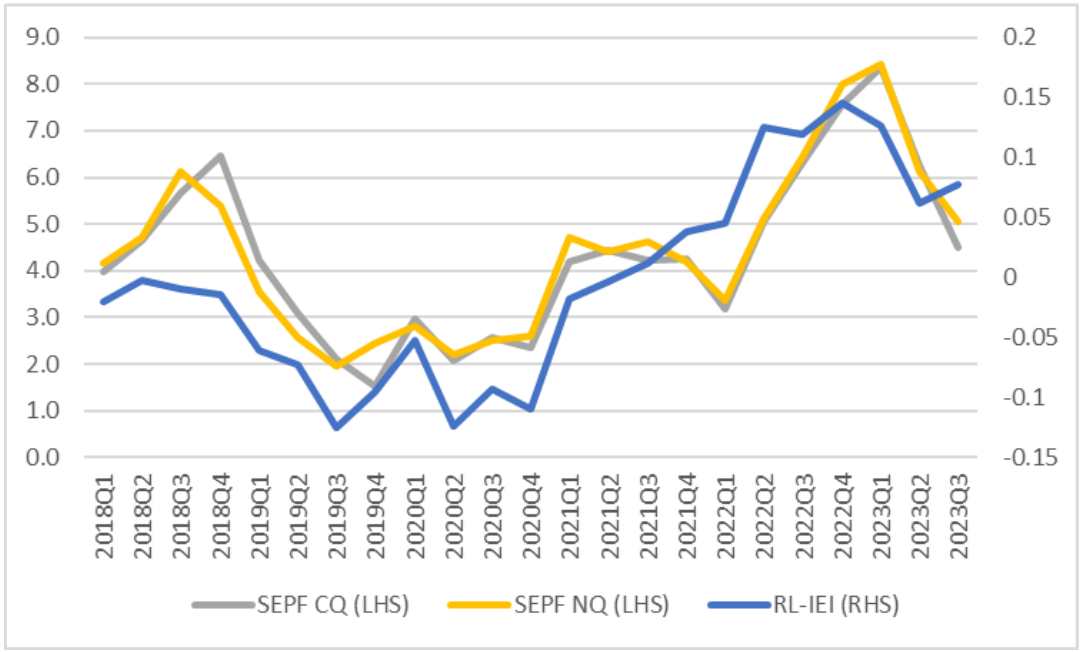
* *Period coverage: 2018 Q1 – 2023 Q3*

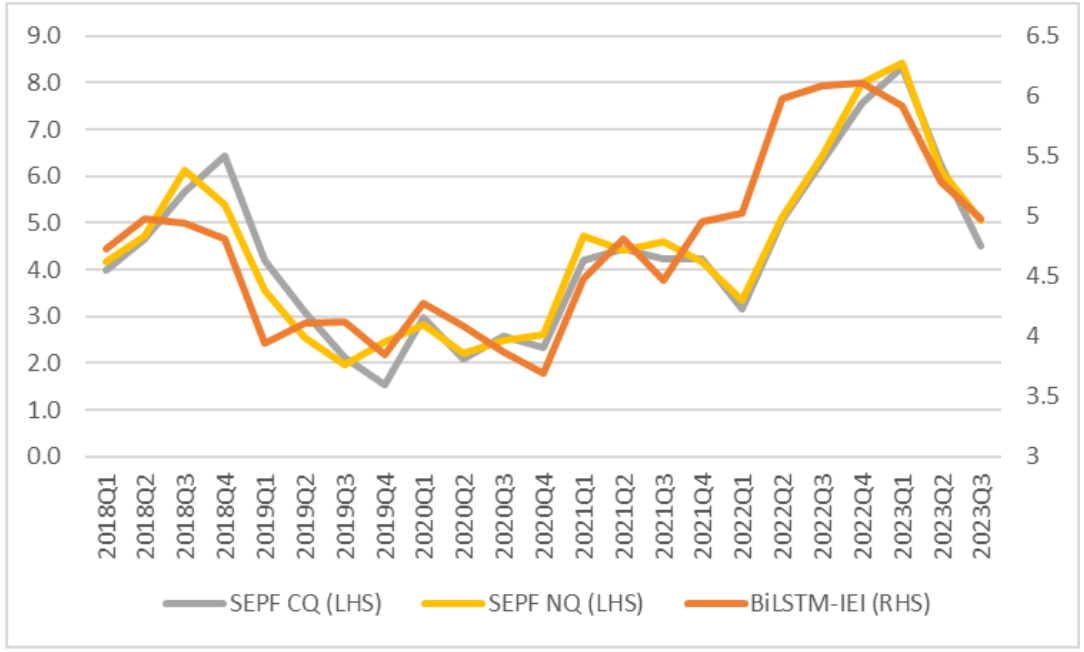
** *Period coverage: 2020Q1 – 2023 Q3*

Note: Q2 2020 is not included in the sample as there were no survey results during this quarter.

Source: Authors' estimates

INIs vs Professional Forecasters' Expectations



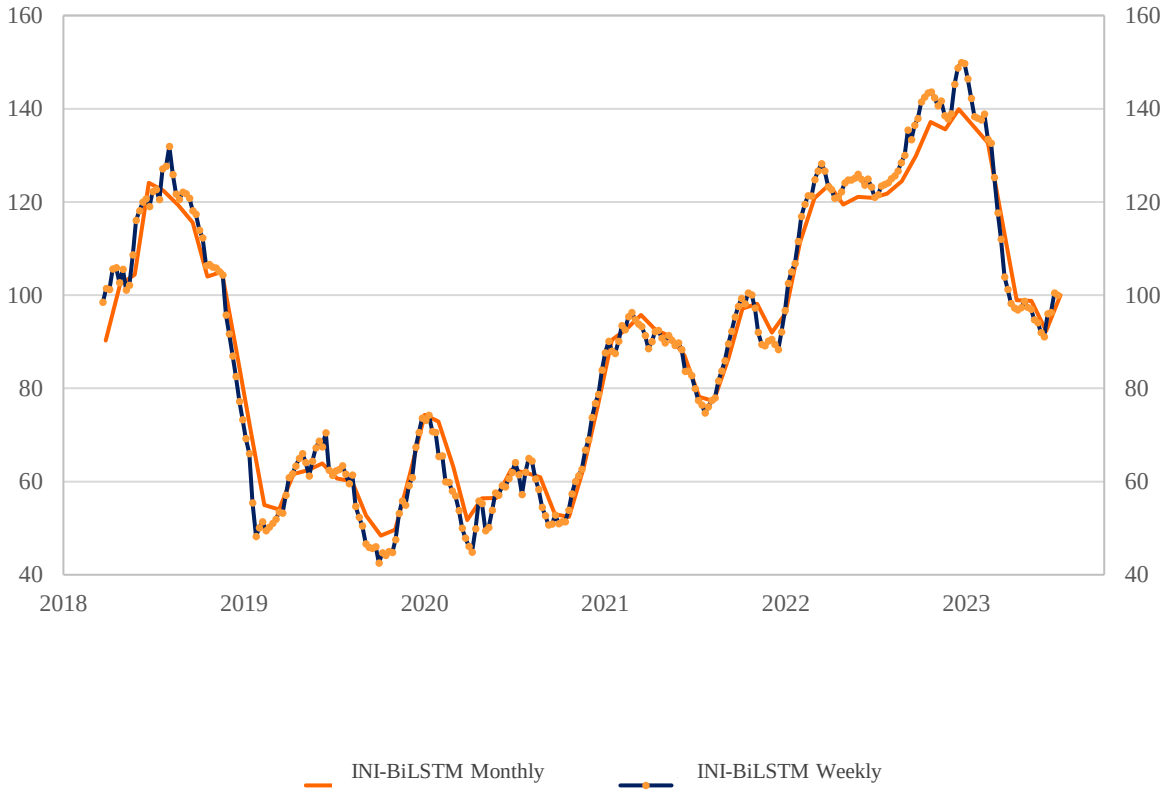


Correlation between INIs and BSEF

	SEPF-CQ	SEPF-NQ
	Overall (2018Q1 - 2023Q3)	
RL lexicon	0.82	0.86
BiLSTM(32)	0.84	0.87

Policy Relevance and Recommendation

Weekly and Monthly tracker from IEI



Example of a Wordcloud based on labelled news articles

Highlights of the Study

We construct an inflation news index (INI) using text analysis on news articles

- The INI was generated using two methods: (1) Dictionary-based method with Reinforcement Learning and (2) Machine learning-based method using Bidirectional Long Short-term Memory (Bi-LSTM)

The INI shows strong correlation with actual inflation and inflation expectations

- The resulting indices all show strong correlation with the actual inflation and inflation expectations (BSEF) for the month with the potential to be used as 1-month ahead leading indicators of inflation

The INI has the potential to be used for inflation forecasting

- The potential of the INI as a leading indicator to inflation can be used for forecasting and nowcasting exercises.

Hot off the press: An inflation news index for the Philippines using reinforced lexicons and memory-based deep learning

Cherrie R. Mapa, Ma. Ellysah Joy T. Guliman,
Alan Chester T. Arcin, Jacqueline Margaux G. Herbo, Sanjeev A. Parmanand and Genna Paola S. Centeno

Bangko Sentral ng Pilipinas

CBSL International Research Conference

13 December 2024

**The views expressed in this paper are those of the authors and do not necessarily reflect those of the BSP. Any errors and omissions are the sole responsibility of the authors.*