

# Classifying Policy Issue Frame Bias in Philippine Online News<sup>\*</sup>

Dela Cruz, Jose Mari Luis M.<sup>1</sup>[0009–0009–8100–2991] and Estuar, Maria Regina Justina E.<sup>1</sup>[0000–0003–0767–9485]

Department of Information Systems and Computer Science,  
Ateneo de Manila University, Quezon City 1108, Philippines  
`mari.delacruz@obf.ateneo.edu`; `restuar@ateneo.edu`

**Abstract.** Media plays an important role in disseminating news to the public [1,14]. However, the selection of what information to write about, how information is presented, and when it is broadcasted are within the control of the media outlet. Media can therefore shape the consumers' opinion based on how publicly available news [1,9] contents are published. Framing bias occurs when the author selects and highlights aspects of the news [8]. This study developed a model to classify the policy issue frames in select Philippine online news articles. Media Frame Corpus [5] was tested on the classification of policy issue frames using supervised learning methods including Bidirectional Encoder Representations from Transformers (BERT), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), linear support vector machine (SVM), and logistic regression (LR). Results showed that the BERT model performed best with an accuracy of 74.73% with political frame (Frame 13) and economic frame (Frame 1) as leading policy issue frames. Implementing the MFC-BERT classification model on the Philippine dataset shows that there is a dominant policy issue frame across all selected media outlets. However, there is a significant difference in the usage of policy issue frames among these media outlets except for the Fairness and Equality frame (Frame 4) and the Quality of Life frame (Frame 10). Initially, ambiguous topics among media outlets when observed over time exhibit noticeable policy issue framing bias. However, a gradual transition to having similar policy issue frames among most media outlets occurred to these topics, similar to the unambiguous topics.

**Keywords:** Framing · Policy Issue Frames · Philippine Online News Media · Text Classification.

## 1 Introduction

Media circulates news such as political affairs, economic circumstances, and other issues around the globe [1, 14]. News dissemination comes in many forms and sources including print media, broadcast media, and online media. The public

---

<sup>\*</sup> Ateneo de Manila University - Social Computing Laboratory of Ateneo

depends on news articles as one of the primary sources of information source [1,9]. With the advent of the Internet and the emergence of different online platforms, news articles are now available online. These online articles are now considered vital information sources that extend traditional media sources and compete with new sources like social media [7,9,14].

The media’s most integral obligation in society is to communicate vital information and events to the people [10,13,15], dedicated to having informed citizenship [15]. However, different news media outlets have different ways of reporting the news. Given that news articles have a vital function in forming individuals’ opinions [1,9], the variations in narratives on how the media presents an issue influences the consumers’ recognition and understanding of the event [13]. This process of selecting and highlighting aspects of news is called framing, and the actual presentation of the news is called a frame [8].

Studies on policy issue frames, also known as issue definitions, have been a topic of research for over a decade [2,3]. Policy issues are usually one of the more contentious topics in news writing. Policy issues are defined as economic frames, morality frames, political frames, and quality of life frames [5,6] to name a few. These frames are used by politicians, the media, and the voting public to communicate policy issues [3]. The Policy Frames Codebook [4] is a non-issue specific framing schema, which cuts across all policy issues providing a system for categorizing framing cues across policy issues [3]. Media Frame Corpus (MFC) is an established dataset composed of 39,395 news articles primarily from Western News articles on various policy issues such as tobacco, same-sex marriage, immigration, gun control, and the death penalty, annotated in media framing that used the Policy Frame Codebook as a guide [5].

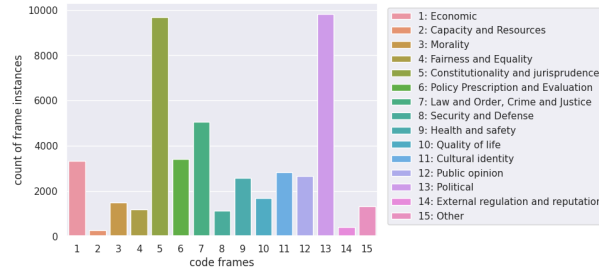
This study aims to test the feasibility of developing a policy issue frame classification model from the Media Frame Corpus that can be applied to Philippine online news.

## 2 Methodology

Bidirectional Encoder Representations from Transformers (BERT), Gated Recurrent Unit (GRU), Long short-term memory (LSTM), Linear Support Vector Machine (SVM), and Logistic Regression (LR) models were used on Media Frame Corpus (MFC) for policy issue frame classification. The best-performing model was used to classify policy issue frames using the dataset extracted from Philippine online news articles.

### 2.1 Selection of Media Frame Corpus (MFC) Training Dataset

MFC contains annotations of policy issue framing on various issues. The MFC dataset was narrowed down to sentences and phrases where agreement was found with at least two annotators. The final dataset resulted in a total of 46799 rows of sentences or phrases. The distribution of the 15 policy issue frames used in MFC is shown in Figure 1.



**Fig. 1.** Distribution of the different policy issue framing - code frames found in the MFC.

## 2.2 Selection of The Philippine Online News Dataset

GMA, Rappler, Manila Bulletin, Philippine News Agency (PNA), and Sunstar Philippines were used to represent Philippine online news. Selection criteria for online news included popularity, use of the English language, and availability of extracting news articles for a longer time period. The final Philippine dataset comprised of headlines and leads of the online news hailing from the different media outlets. A total of 21603 most recent news from online national news articles written in English were extracted from May 5, 2023, to May 7, 2023, using the first 2000 news articles from each media outlet.

Beautifulsoup and selenium packages were used for the extraction of online news articles. Standard data extraction methods were used. A Chrome driver was used to visit the media outlets' pages for national headline news. The HTML elements tags that lead to relevant article links and metadata (e.g. <p>, <li>) were identified for each media outlet. Using these identifiers, each news' URL, author, headline, and date were collected. The Chrome driver was used to scrape the metadata and the content of the article using predetermined HTML element tags of each media outlet from this list.

A dataframe is composed of rows of news articles, including metadata such as the source, headline, and date. The standard method including removing links, photo captions and references, and media outlets' specific terms (e.g. media outlet site name), and the separation of each paragraph of news articles into respective rows were used for pre-processing.

Four variations of the dataset were developed. The first dataset comprised of news articles from all media outlets. The second dataset comprised of news articles separated per media outlet. Criteria for the third dataset, comprised of nonambiguous articles narrowed down using publications from '2023-04-23' to '2023-04-29', and selected using top keywords: 'Sudan', 'Balikatan', 'SIM card registration'/'Sim registration. The fourth dataset comprised of ambiguous articles using the following keywords, 'COVID'/'Covid', 'DepEd', and 'China'.

### 2.3 Developing Classification Models

Bidirectional Encoder Representations from Transformers (BERT), Gated Recurrent Unit (GRU), Long short-term memory (LSTM), and supervised machine learning classifiers (Linear Support Vector Machine and Logistic Regression) were used in training the model on the MFC dataset. Tokenization was used for BERT. Word embeddings were used for LSTM and GRU. term frequency-inverse document frequency (tf-idf) was used for Linear SVM and logistic regression. All models used 85-15 train-test split, stratified according to the 15 frames.

### 2.4 Classification of Policy Issue Frames on Philippine Dataset using BERT

The resulting trained model on BERT was used to classify policy issue frames on all variations of the Philippine dataset, where each sentence or phrase is limited to only one policy issue frame. Accuracy score was used to test the performance of the model.

### 2.5 Testing the Difference of Policy Frames among Media Outlet

Chi-square goodness of fit was used to test for the significance of differences in the use of policy issue frames labeled per article among the five media outlets. Actual frame counts were used for observed values while the mean score of observed values was used for expected values.

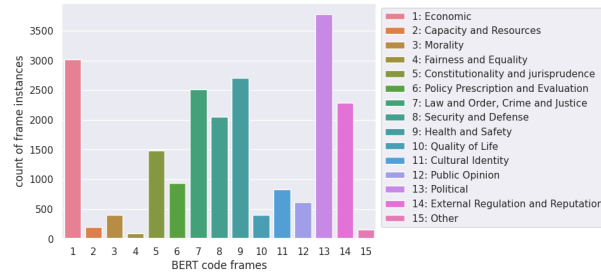
## 3 Results and Discussion

Accuracy scores resulting from the MFC training for all the classification methods are shown in Table 1. Results showed the accuracy scores, specifically for the BERT, LSTM, and GRU are similar to reference studies that were also trained using MFC [11, 12] validating the use of MFC corpus for establishing ground truth.

**Table 1.** Performance Metric of the Models trained with MFC

Model	Accuracy(%)	Referenced Studies' Accuracy
BERT	74.73	71.50
LSTM	70.78	52.10
GRU	70.74	53.70
SVM	63.07	N/A
LR	68.25	N/A

BERT model was the highest performing model, followed by LSTM, and then GRU. On the other hand, the SVM and LR were the least accurate.

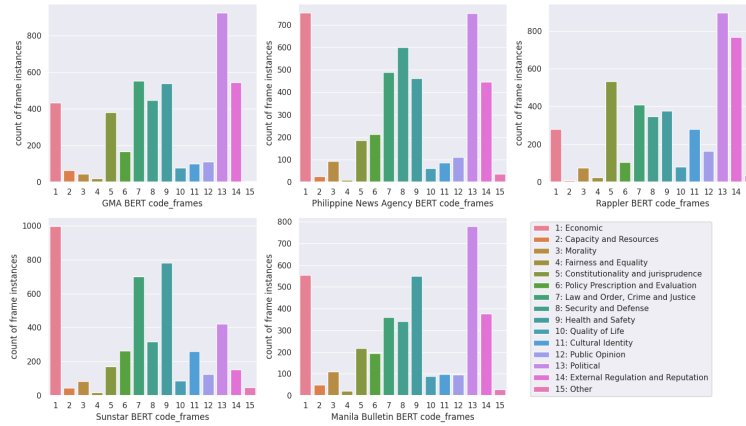


**Fig. 2.** Distribution of the policy issue frames on the aggregated dataset using BERT Model

### 3.1 Frame Classification on the Aggregated Philippine Dataset

Figure 2 shows the frequency count of resulting frame classification when BERT model was implemented on the Philippine dataset. The political frame (Frame 13) had the highest count followed by the economic frame (Frame 1). The next dominant frame is the health and safety frame (Frame 9).

### 3.2 Comparison of Frame Classification per Media Outlet



**Fig. 3.** Distribution of the policy issue frames on the dataset per Media Outlet using BERT Model

Figure 3 shows the frequency count of frame instances of each policy issue frame in each media outlet. Specifically for the BERT model, when the dataset is divided into different media outlets the general observation is that the prevalent frames vary for each media outlet. The political frame (Frame 13) is the

frame with the highest count from most outlets however, the distribution of the remaining frames varies per outlet. Furthermore, the test for goodness of fit showed that there is a significant difference among preferences on policy issue frame per media outlet for all frames [  $p < 0.05$ ] except for the Fairness and Equality frame (Frame 4) [ $p=0.14$ ] and the Quality of Life frame (Frame 10) [ $p = 0.15$ ].

### 3.3 Comparison of Frame Classification per Media Outlet on Unambiguous Topics

**Table 2.** Comparison of Top 3 Policy Issue Frames where topic = Sudan.

	Sudan Top Frame		
	Top 1	Top 2	Top 3
<b>GMA</b>	Economic	Security and Defense	Political
<b>PNA</b>	Security and Defense	External Regulation and Reputation	Health and Safety
<b>Rappler</b>	Security and Defense	Quality of Life	Capacity and Resources/External Regulation and Reputation
<b>Sunstar</b>	Security and Defense	Economic	External Regulation and Reputation
<b>Manila Bulletin</b>	Security and Defense	External Regulation and Reputation	Economic

**Table 3.** Comparison of Top 3 Policy Issue Frames where topic = Balikpapan.

	Balikatan Frame		
	Top 1	Top 2	Top 3
<b>GMA</b>	Security and Defense	External Regulation and Reputation	Cultural Identity/Law and Order, Crime and Justice
<b>PNA</b>	Security and Defense	External Regulation and Reputation	Cultural Identity
<b>Rappler</b>	Security and Defense	External Regulation and Reputation	Political
<b>Sunstar</b>	Security and Defense	External Regulation and Reputation	Cultural Identity
<b>Manila Bulletin</b>	Security and Defense	External Regulation and Reputation	Cultural Identity

**Table 4.** Comparison of Top 3 Policy Issue Frames where topic = SIM Registration.

	SIM Registration Top Frame		
	Top 1	Top 2	Top 3
<b>GMA</b>	Policy Prescription and Evaluation	Political	Constitutionality and Jurisprudence
<b>PNA</b>	Policy Prescription and Evaluation	Economic	Political
<b>Rappler</b>	Constitutionality and Jurisprudence	Policy Prescription and Evaluation	Political
<b>Sunstar</b>	Policy Prescription and Evaluation	Economic	Constitutionality and Jurisprudence
<b>Manila Bulletin</b>	Policy Prescription and Evaluation	Economic	Political

Selected topics include the Repatriation of Filipinos from Sudan, the Balikpapan Exercise between the United States and the Philippines, and Sim Card Registration.

**Frames per Issue:** Tables 2,3, and 4 show the top 3 frames with the highest count per media outlet. The Sudan topic, illustrated in Table 2, is about the repatriation of Filipino from strife-torn Sudan and the actions of the Philippine

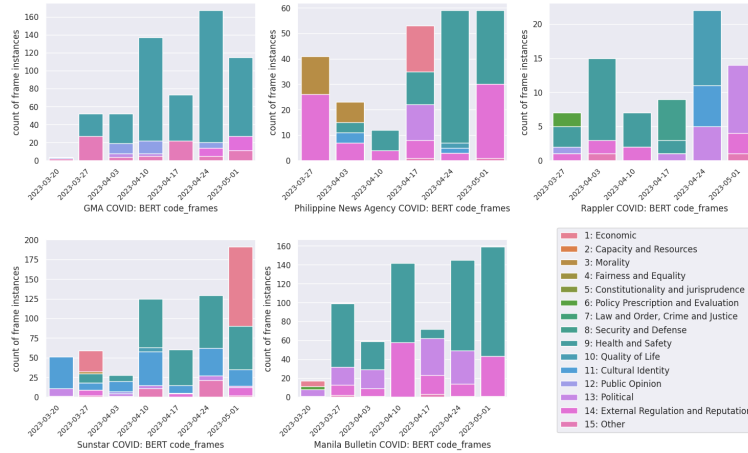
government in helping safeguard the welfare of the Filipinos. All media outlets, except GMA, used the security and defense frame (Frame 8) as the top frame for the issue. GMA, however, used the economic frame (Frame 1) as its number one frame, followed by the security and defense frame (Frame 8). It is observable, however, that media outlets do not have the same succeeding frames.

The Balikatan topic, depicted by Table 3, is about the annual joint military exercise between the United States Army and the Philippine Army. It is observable that the security and defense frame (Frame 8) is the consensus top 1 frame for all media outlets with external regulation and reputation frame (Frame 14) as the top 2 frame. However, the top 3 frame differs for each outlet.

The SIM topic in Table 4 is about how the SIM registration ACT is being implemented in the Philippines and the challenges faced by policymakers as well as the public. The policy prescription and evaluation frame (Frame 6) is the dominant frame for all outlets except Rappler, which has the constitutionality and jurisprudence frame (Frame 5) as its top frame. Similarly, the succeeding secondary and tertiary frames vary per media outlet. From these observations, we can infer that each topic has a primary frame.

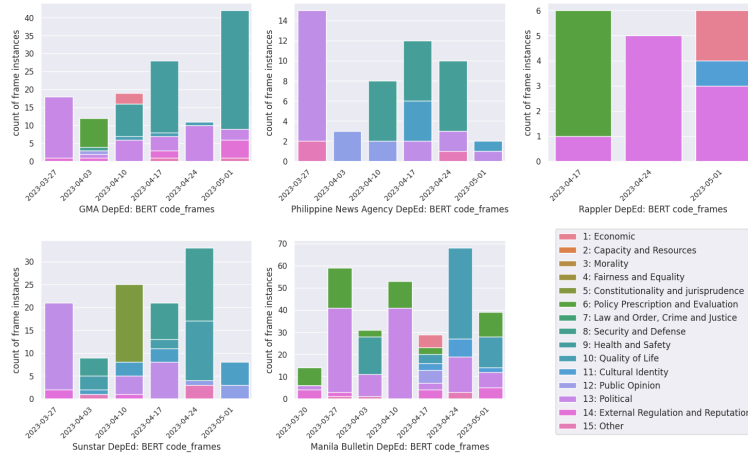
### 3.4 Comparison of Frame Classification per Media Outlet on Ambiguous Topic

Topics were selected to test if policy issue frame bias will be present when topics are ambiguous. Figure 4, 5, and 6 show the policy issue frame count distribution per Media Outlet from March of 2023 to May of 2023.



**Fig. 4.** Distribution of Policy issue frames over time per Media Outlet on keyword = COVID.

Figure 4 shows at first glance that the dominant frame is Health and Safety frame (Frame 9). However, it is also observed that there are changes in the use of policy issue frames over time which varies across media outlets. For example, there was a shift from Health and Safety frame (Frame 9) to the Economic frame (Frame 1) in Sunstar. Additionally, Rappler shifted its dominant frame to Political frames (Frame 13) from the Health and Safety frame (Frame 9).



**Fig. 5.** Distribution of Policy issue frames over time per Media Outlet on keyword = DepEd.

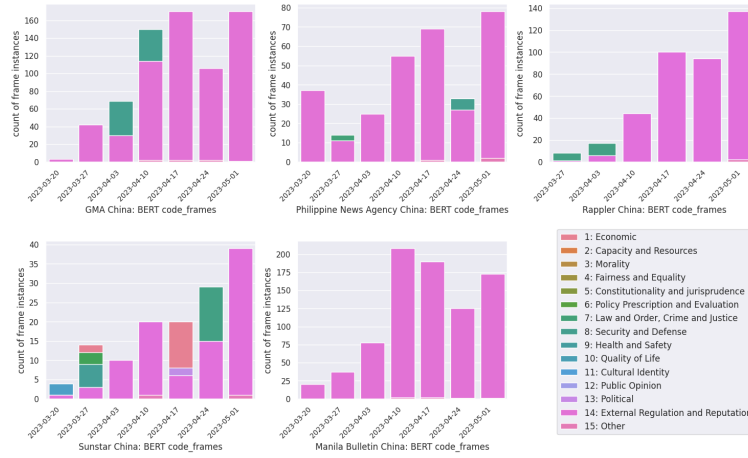
Figure 5 shows that only three out of the five media outlets begin with the Political frame (Frame 13) and transition to different frames over time. Rappler, on the other hand, with limited stories on DepEd started with Policy Prescription and Evaluation frame (Frame 6) as its initial frame, then transitioned to the Political frame (Frame 13) as its final frame. Manila Bulletin shows two policy issue shifts from Policy Prescription and Evaluation frame (Frame 6), followed by the Political frame (Frame 13), then ended with the Cultural Identity frame (Frame 11).

On the other hand, Figure 6 shows one dominant policy issue frame over time namely, the External Regulation and Reputation frame (Frame 14).

## 4 Conclusion

The Media Frame Corpus served as ground truth to predict policy issue frames for the Philippine online news dataset. Albeit trained on a Western-based corpus, the MFC-BERT classification model is applicable to the Philippine dataset. When all articles are combined as a whole regardless of media outlets, a dominant policy issue frame becomes evident. There is a significant difference in the use





**Fig. 6.** Distribution of Policy issue frames over time per Media Outlet on keyword = China.

of policy issue frames among media outlets except for the Fairness and Equality frame (Frame 4) and the Quality of Life frame (10). However, similar policy issue frames are observed when unambiguous topics are chosen as an additional criterion. For ambiguous topics though, policy issue frame bias is noticeable among media outlets at the onset. Policy issue frame shift towards similar usage over time across most media outlets.

## References

1. Al-Hindawi, F.H., Ali, A.H.: A pragmatic study of cnn and bbc news headlines covering the syrian conflict. *Advances in Language and Literary Studies* **9**(3), 43–51 (2018)
2. Bogaerts, J., Carpentier, N.: The postmodern challenge to journalism: Strategies for constructing a trustworthy identity. In: *Rethinking Journalism*, pp. 60–71. Routledge (2013)
3. Boydston, A.E., Card, D., Gross, J., Resnick, P., Smith, N.A.: Tracking the development of media frames within and across policy issues (2014)
4. Boydston, A.E., Gross, J.H., Resnik, P., Smith, N.A.: Identifying media frames and frame dynamics within and across policy issues. In: *New Directions in Analyzing Text as Data Workshop*, London (2013)
5. Card, D., Boydston, A.E., Gross, J.H., Resnik, P., Smith, N.A.: The media frames corpus: Annotations of frames across issues. In: *Proceedings of ACL* (2015), <https://www.aclweb.org/anthology/P15-2072/>
6. Card, D., Gross, J.H., Boydston, A., Smith, N.A.: Analyzing framing through the casts of characters in the news. In: *Proceedings of the 2016 conference on empirical methods in natural language processing*. pp. 1410–1420 (2016)
7. Dallmann, A., Lemmerich, F., Zoller, D., Hotho, A.: Media bias in german online newspapers. In: *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. pp. 133–137 (2015)

8. Entman, R.M.: Framing: Towards clarification of a fractured paradigm. *McQuail's reader in mass communication theory* **390**, 397 (1993)
9. Hamborg, F., Donnay, K., Gipp, B.: Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries* **20**(4), 391–415 (2019)
10. Jenkins, J., Nielsen, R.K.: Proximity, public service, and popularity: A comparative study of how local journalists view quality news. *Journalism Studies* **21**(2), 236–253 (2020)
11. Liu, S., Guo, L., Mays, K., Betke, M., Wijaya, D.T.: Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence. In: *Proceedings of the 23rd conference on computational natural language learning (CoNLL)* (2019)
12. Naderi, N., Hirst, G.: Classifying frames at the sentence level in news articles. *Policy* **9**, 4–233 (2017)
13. Pristianita, S., Marta, R.F., Amanda, M., Widiyanto, Y.N., Boer, R.F.: Comparative analysis of online news content objectivity on covid-19 between detik. com and Kompas. com. *Informatologia* **53** (2020)
14. Spinde, T., Rudnitskaia, L., Mitrović, J., Hamborg, F., Granitzer, M., Gipp, B., Donnay, K.: Automated identification of bias inducing words in news articles using linguistic and context-oriented features (2021). <https://doi.org/https://doi.org/10.1016/j.ipm.2021.102505>, <https://www.sciencedirect.com/science/article/pii/S0306457321000157>
15. Tiffen, R., Jones, P.K., Rowe, D., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Mazzoleni, G., Papathanassopoulos, S., et al.: Sources in the news: A comparative study. *Journalism studies* **15**(4), 374–391 (2014)