Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 3, 2025
DOI: 10.30829/zero.v9i3.26837
Page: 980-990

Text Mining and News Sentiment Analysis of the PPRT (Domestic Worker Protection) Bill in Three Online News Media From 2004 to 2024

[1] Dian Novita Kristiyani
Department of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia

[2] Richrad Gordon Mayopu
Department of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia

Article Info

Article history:

Accepted, 26 December 2025

Keywords:

Domestic Workers;
Online Mass Media;
PPRT Bill;
Sentiment Analysis;
Text Mining.

ABSTRACT

The Domestic Worker Protection Bill (RUU PPRT) has been a critical issue in Indonesia, yet its legislative process has stagnated for two decades, leading to intense public discourse. This study aims to analyze the sentiment and narrative dynamics of RUU PPRT news coverage in online media, as well as the media's role in shaping public opinion. Employing a Text Mining and Lexicon-Based Sentiment Analysis approach, enhanced with adaptations for the Indonesian lexicon, this research analyzes 387 news articles from three prominent online media outlets (Tempo, Kompas, and VOA News) published between 2004 and 2024. The findings reveal that positive sentiment dominates with 58.1%, followed by negative sentiment at 31.3%, and neutral sentiment at 10.6%. Tempo was identified as the most active media outlet covering this issue. These results indicate that the mass media plays a significant role in shaping the pattern of public discourse regarding the PPRT Bill, particularly through the dominance of positive sentiment in its reporting and confirm that lexicon-based sentiment analysis can systematically capture the dynamics of complex socio-political narratives.

Corresponding Author:

Dian Novita Kristiyani,
Department of Information Tecnology,
Satya Wacana Christian University, Salatiga, Indonesia
Email: dian.kristiyani@uksw.edu

1. INTRODUCTION

Domestic workers (PRT) in Indonesia are among the most vulnerable labor groups, frequently enduring exploitation, violence, and a severe lack of legal protection. Millions of PRT operate without adequate legal frameworks, rendering them susceptible to various human rights violations. This precarious situation leads many PRT to feel devoid of rights, akin to being "slaves" unable to free themselves[1]. The urgency of this issue is underscored by alarming statistics: the National Network for Domestic Advocacy (JALA PRT) reported 3,308 cases of violence against domestic workers from 2021 to February 2024 [2], while National Commission on Violence Against Women documented a continuous increase, reaching 2,344 cases from 2005 to 2022. [3]. These injustices encompass unfair wages, exploitative working hours, lack of social security, and both verbal and physical abuse from employers, exacerbated by the absence of specific protective legal regulations.

Despite their significant, albeit indirect, economic contribution to both household and national economies[4], PRT remain marginalized. This aligns with Nancy Fraser's (2016) analysis in Contradictions of Capital and Care, where capitalism's dichotomy between production and reproductive work often positions women, particularly in the informal sector like domestic work, in subordinate, unrecognized, and unprotected roles, making them vulnerable to exploitation [5]. The root cause of this systemic violence against female and child domestic workers is the Rechtvacum a regulatory void concerning their legal protection in Indonesia. Under the existing Labor Law, the domestic worker profession is not recognized, leaving them without legal safeguards, decent working conditions, social security, or labor rights [6].

The Domestic Worker Protection Bill (RUU PPRT) was initiated as a crucial legislative effort to rectify this, aiming to guarantee rights and improve working conditions for PRT. However, compared to other bills, the RUU PPRT has faced an unprecedented 24-year legislative stagnation, enduring four changes in the House of Representatives (DPR) periods without ratification [7]. This prolonged debate has triggered significant responses from civil society, the public, and mass media. In this protracted socio-political context, mass media plays a vital role as an agent of public communication, shaping public opinion and influencing policy direction [8]. A careful analysis of media reports can reveal how the RUU PPRT issue is framed and the dominant sentiment in public narratives, as media can frame issues positively, neutrally, or negatively, significantly impacting public perception and policymaking [9].

Sentiment analysis, an information mining technique, automatically processes textual data to identify an author's viewpoint, typically categorizing it as positive, negative, or neutral [10][11]. This approach is crucial for understanding how the RUU PPRT narrative is formed, disseminated, and received by the public, especially in the digital era where online media's accessibility greatly influences perception. Natural Language Processing (NLP), a branch of artificial intelligence, enables computers to process, understand, and classify text [12]. While various studies have analyzed public policy and media reporting on the RUU PPRT using qualitative and quantitative methods, research employing explicit text mining, particularly methods like Latent Semantic Analysis (LSA), to analyze the RUU PPRT discourse remains minimal [13].

From a computational and applied-mathematics perspective, systematic sentiment quantification over large-scale longitudinal corpora requires formal weighting models and polarity aggregation functions that enable reproducible and scalable discourse measurement. Integrating TF-IDF term weighting with lexicon-based sentiment scoring provides a mathematical representation of discourse salience and polarity, allowing consistent comparison of narrative dynamics across time and heterogeneous media sources.

This study aims to fill this gap by applying a computational sentiment-analysis framework on a large-scale corpus of news articles from 2004 to 2024, integrating TF-IDF weighting and lexicon-based polarity scoring to model discourse salience and sentiment dynamics. Our research seeks to: (1) determine the distribution of sentiment (positive, negative, neutral) in RUU PPRT news coverage across Tempo, Kompas, and VOA News; (2) analyze the evolution of sentiment dynamics over time and identify key influencing events; and (3) explore differences in sentiment patterns and narratives among the three media outlets. This analysis is expected to provide both substantive socio-political insights and a replicable computational framework for measuring long-term policy discourse dynamics, contributing to critical media literacy and evidence-based public policy analysis.

2. RESEARCH METHOD

This study seeks to provide a better picture of media reporting to the online masses from 2004 to 2024. The researcher uses a text mining approach to answer the above problem by identifying the main topics that developed from the discourse of the PPRT Bill. Youthfulness maps the leading actors that emerge reported by the media and relates the relationship between various elements in the special news about the PPRT Bill. Next, measure sentiment tendencies (positive, negative, or neutral) in the news of the PPRT Bill.

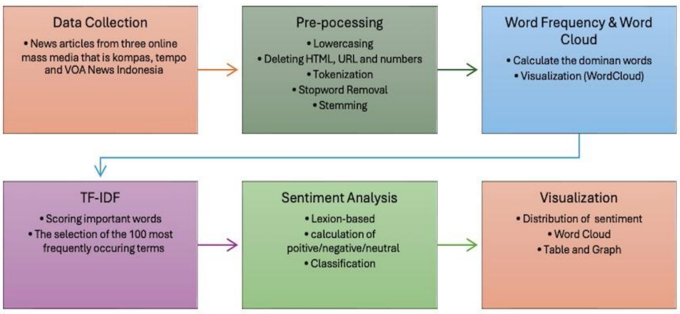

Figure 1. Research Stages

The stages of this research are shown in Figure 1. The following is a description of the stages in the study that has been carried out: Data Collection. The data used in this study consists of online news articles discussing the Domestic Worker Protection Bill (RUU PPRT). These articles were collected from three leading online media outlets in Indonesia: Tempo (tempo.co), Kompas (kompas.com), and VOA News (voaindonesia.com). The selection of these three media was based on their reputation as credible news sources with a broad readership in Indonesia. The data collection timeframe spans from January 2004 to December 2024, covering the entire duration of the RUU PPRT discussion to date and consist to 390 news articles.

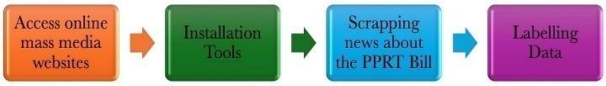

Figure 2. Data Scraping Stages

Figure 2 shows the stages of data scraping. Web Scraping allows authors to quickly and easily retrieve many news articles, facilitating more in-depth investigation and analysis of data [14]. The data in this study was obtained by using the web scraping method to collect news articles from Tempo, kompas.com, detik.com, between news, and VOA Indonesia. Research conducted by Ahmed [15] states that the web scraping method is proven effective in collecting structured text data for scientific analysis.

The first is to access the three online mass media websites to get all news related to the PPRT Bill (RUU PPRT) from 2004 to 2024. This data is the primary source of news related to the PPRT Bill. This stage aims to identify and collect URLs and relevant news. Furthermore, Data Scraping is carried out using scraping tools. This is done to prepare the working environment required in the data scraping process. Then, the news scraping of the PPRT Bill is carried out, using a web scraping process to automatically retrieve all news from the required online media web pages. The data collected from each three online media includes title, date, year of publication, news content, and URL link. After the data is collected, data labelling is carried out to classify news according to its category, such as based on sentiments, topics, or attitudes towards the PPRT Bill. The total scraping data obtained was 445 news articles. Then, pre-processing was done to remove duplicate data from 3 online mass media outlets and convert it into 387 news articles.

Table 1. Dataset Classification

Article (s)	Number (s)
Positive	225
Negative	121
Neutral	41
Total	387

The table above shows the classification, visualization was carried out to see how much the difference in sentiment for each article from the three media was obtained.

2.1 Text Pre-Processing

Text pre-processing is one of the critical stages in sentiment analysis [16]. This process is carried out to eliminate noise in the data. First, duplicate news is removed to improve the validity of the dataset. The second is lowercase, which aims to convert text data to lowercase letters and eliminate characters other than the letters a to z. Then, the third step is to remove punctuation and spacing in news content data, and the fourth step is tokenization. This stage also omits certain characters, such as punctuation marks, and filters by text length [17]. To get appropriate and meaningful words that are further processed in classification, filtering using stop wards (eliminating unimportant/meaningless words) is necessary. Stop ward removal and tokenization were the pre-processing techniques used in the system, along with TF-IDF as the vectorization technique [18]. In this study, TF-IDF is employed not as a supervised classification algorithm, but as a term-weighting and salience representation technique. References to Support Vector Machine (SVM) are omitted, as sentiment classification is lexicon-based. From this stop word stage, sentiment analysis will be carried out to focus on words that have important meanings. And the process continues with stemming, which changes the word into the form of the root word. From this process, a word cloud will be generated. To classify data, it is necessary to transform it [19] into a form that a computer can process. Weighting or giving values is carried out to carry out a transformation. In previous research using the SVM method, TF-IDF was the weighting that produced the best accuracy [20]. TF-IDF, an algorithm used in NLP, assesses the significance of a term within a document compared to a larger corpus of documents [21]. In this research, weighting is carried out using TF-IDF, which transforms words into vector form to be processed using the Support Vector Machine method. The TF-IDF steps and formula are below [14], [22].

Step 1: Tokenization
Text was tokenized using Python's NLTK library with unigram segmentation.
Step 2: Data Cleaning
Stop words, and irrelevant symbols were removed to ensure clean data.
Step 3: Lemmatization
Words were normalized to their root forms (e.g., "running" and "ran" became "run").
Step 4: Word Frequency Counting
Words occurring fewer than five times were excluded from analysis.
Step 5: Constructing a Term–Document Matrix (TDM)

A term document matrix was built using TF–IDF weighting as shown in Equation (1).

$$TF - IDF = TF(t_i, d_j) \times \log\left(\frac{N}{N(t_i)}\right) \tag{1}$$

Where t_(i) is the ith term; d_i represents the jth document; N is the total number of all documents; N (t_(i)) denotes the number of documents which contain t_(i) features.

The sentiment lexicon was developed through a corpus-driven qualitative procedure. Lexical candidates were identified by repeatedly reading and mapping frequently occurring evaluative expressions in RUU PPRT news articles. Each term was then examined within its surrounding textual context to ensure consistent polarity orientation (positive, negative, or neutral) in public policy discourse. Only terms that appeared recurrently across different news sources and demonstrated stable contextual meaning were retained in the final lexicon. This contextual validation procedure ensures that the lexicon reflects discourse-specific sentiment expressions rather than general-purpose affective words.

After pre-processing, the process continues with the use of a lexicon.

We use a lexicon model to analyze the sentiment, and the detailed steps are shown as follows.

$$S(d) = \sum i = 1NL(wi) \tag{2}$$

Where each token's polarity L(wi)L(w_i) is +1 for positive words, −1 for negative words, and 0 otherwise.
The final label Y(d)Y(d) follows:

$$Y(d) = \{positive, S(d) > 0\ [4pt]negative, S(d) < 0\ [4pt]neutral, S(d) = 0$$

Where,
d: the preprocessed document (list of tokens).
w_i:the ith token in d,with i=1,2,...,N.
N: total number of tokens in d.
L(w): lexicon mapping
+1 if w belongs to the positive list
-1 if w belongs to the negative list
0 otherwise [23][24]

2.2 Sentiment Labelling

The sentiment labelling approach utilises a lexicon-based method, where each word in a document is assigned a polarity score according to its presence in predefined positive, negative, or neutral lists. The overall sentiment of the document is determined by summing these scores, resulting in a positive, negative, or neutral classification. This method enables systematic and automated sentiment analysis tailored to the specific contextual vocabulary.

Table 2. Sentiment Analysis

Sentiment Category	Words Included
Positive Words	Protect, agree, support, fair, important, recognized, benefit, worthy, struggle, discuss, legal, propose, approve, form, urge
Negative Words	Reject, ignore, problem, threat, violate, bad, no, discrimination, exploitation, violence, delay, hold, valid, harsh, disappointed, cancel, fail, condemn
Neutral Words	All other words not included in the positive or negative lists

The table presents the lexicon used for the automatic sentiment analysis. Words are grouped into three categories based on their predefined polarity: positive, negative, and neutral. The positive and negative lists contain Indonesian terms that explicitly express supportive or critical sentiment within the context of the analysis. Any word not included in these two lists is treated as neutral and does not contribute to the sentiment score. This lexicon serves as the foundation for the rule-based scoring system, where each positive word contributes +1, each negative word contributes −1, and neutral words contribute 0 to the final sentiment classification.

3. RESULT AND ANALYSIS

The debate on the ratification of the PPRT Bill contained various responses from both the public, observers of social movements, and the legislature. This polemic has become widely discussed in the news because it concerns the issue of most citizens working without guarantees and clear legal protection. This problem is increasingly in the spotlight when the discourse on adopting the PPRT Bill is not immediately passed, as the process is over 20 years old. The media is one of the information funnels for the public to explore the dynamics and developments of the ratification of the PPRT Bill. In addition, the media is also a control tool for the House of Representatives and the legislature to see how the response of domestic workers and the public is regarding the slow process of ratification of the PPRT Bill. Many cases occurred, demonstrations for the immediate passage of the PPRT Bill, and several other strong reasons that should encourage the ratification of this bill. In the information era like today, all forms of media speed in providing information in the form of news seem to be a necessity. The press is the main portal in terms of news in the mass media [25]. The media has a responsibility in serving public information. However, sometimes less interesting issues such as the ratification of the PPRT Bill are not newsworthy [7].

An analysis of 387 news articles related to the RUU PPRT demonstrates that the majority of coverage exhibits a positive sentiment, with 225 articles (58.1%) falling into this category. In comparison, negative sentiment is present in 121 articles (31.3%), representing approximately half the proportion of positive articles. Meanwhile, neutral sentiment is observed in just 41 articles (10.6%), making it the smallest segment within the corpus. This distribution suggests that news reporting on the RUU PPRT tends to be predominantly positive, with less frequent negative and neutral perspectives. This distribution indicates that the majority of news coverage tends to have a positive sentiment towards the RUU PPRT. The proportion of negative sentiment is approximately half of the positive sentiment, while neutral sentiment constitutes the smallest portion of the overall corpus.

Twenty years have passed, but until now, during the new presidential administration, the PPRT Bill has not progressed to the PPRT Law. The media continues highlighting this topic, and civil society voices its endorsement. Several press reports have continuously reported this issue over the past few years. In this study, 387 news articles were collected during the pre-processing stage. From these 387 online media news articles, we can see how the sentiment of the PPRT Bill news analysis has been so far.

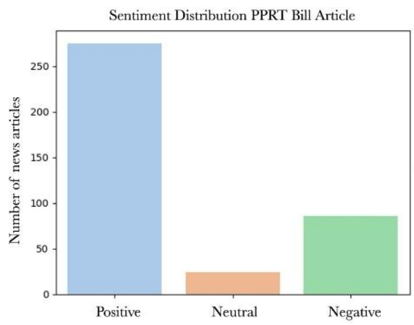

Figure 3. Sentiment Distribution News Article

Figure 3 shows about 387 news articles on the PPRT Bill from 3 online news media, Kompas.com. In Tempo and VOA news, 274 had positive sentiment shown in blue, 86 had negative sentiment shown in green, and orange with a score of 33 had neutral sentiment.Judging from the sentiment analysis of media reports about the PPRT Bill, the majority shows positive sentiment. This led to the ratification of the PPRT Bill, which received much support to be passed immediately.

The amount of positive sentiment towards the discourse on the PPRT Bill news will show how the distribution of sentiment analysis is based on the media.

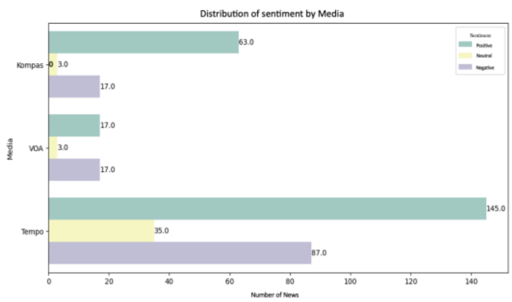

Figure 4. Sentiment distribution based on media

Based on Figure 5, we can see from the three online media that the number of news articles with positive, neutral, or negative sentiments appears to be the highest among the three media, Tempo has the most news on the topic of the PPRT Bill with 267 articles, followed by Kompas with 83 news articles and VOA News with 37 news articles. It shows that all news articles show positive sentiment.

The overall tempo article has 145 news stories with positive sentiments, 35 news with neutral sentiment, and 87 news with negative sentiment. Tempo is the media with the most news reports about the PPRT Bill, with a dominance of positive sentiment. From this, Tempo actively voiced the urgency of the PPRT Bill. The news that emerged also highlighted a lot about the encouragement of civil society, or the steps that should be taken by the House of Representatives and the government.

Kompas also has a relatively large number of news items, with 63 news items having positive sentiment, 3 having neutral sentiment, and 17 having negative sentiment. Meanwhile, VOA between positive and negative sentiment was obtained from 17 news stories, and neutral sentiment from 3 news stories. VOA is more balanced between negative and positive news, with a balanced approach and coverage that delivers its news in a two-sided manner.

The number of news reports and the media's attitude are significant for the primary data. If you look at the chart above, the tempo is very proactive in overseeing the issue of the PPRT Bill. It shows the possibility of partiality towards domestic workers as a vulnerable group. Meanwhile, Kompas is more moderate but still tends to be positive, and VOS provides more balanced coverage in its news.

Positive sentiment can be in the form of support for the bill's immediate passage, which is reflected in the encouragement of activists or the dynamics of the legislation process that have occurred to date. Negative sentiments contain a lot of criticism of the government or the DPR, or the role of political parties, and the increase in cases of violence against domestic workers.

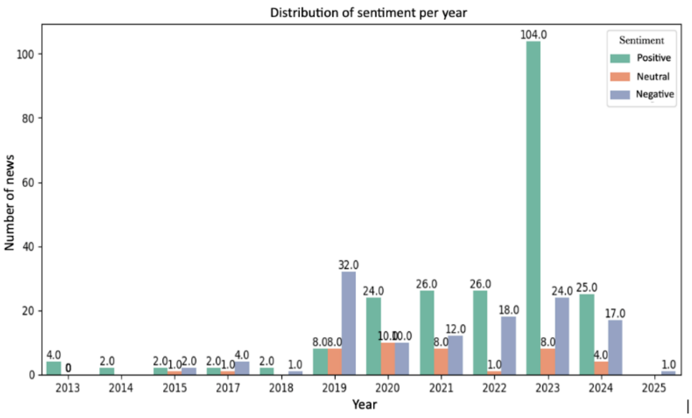

Figure 5. Distribution of news per year

The distribution of sentiment analysis per year is shown in Figure 6. The news of the PPRT Bill can be accessed from 2013 to 2025 from the three Tempo media, Kompas, and VOA news. This is an illustration that for 13 years, the public has been able to obtain information about the topic of the PPRT Bill in online media. Researchers found that 2013 from 2024 to 2012, there had been no news of the PPRT Bill in the three media. This also shows that online media experienced development in 2012. However, from the existing news, we can see that the discourse from many news articles shows the dynamics and development of the ratification of the PPRT Bill over the past 2 decades (20 years).

From the news from 2013 to 2025 above, several graphs are significant to note, namely, in 2019, there are more negative sentiments than positive and neutral sentiments. And in 2023, positive sentiment graphically shows very high compared to neutral and negative sentiment. The year 2019 showed a surge in negative sentiment, most likely triggered by the stagnation of the House of Representatives regarding the ratification of the PPRT Bill, which was proposed for a long time and discussed starting in 2004. Until 2019, it had not been included as a national priority. Many news reports contain criticism of the House of Representatives because it neglects the protection of domestic workers. Civil society protests in 2019 occurred many times and urged the House of Representatives to discuss and pass the PPRT Bill immediately. The media reported the voices of support and disappointment from activist groups, farmers, and domestic workers themselves.

The year 2023 shows the dominance of positive sentiment. In March 2023, the House of Representatives officially approved the PPRT Bill as an initiative bill of the House of Representatives. This is a great victory for the women's and human rights movements. The support of the government and public figures, including President Jokowi and several ministers, expressed open support. The media also welcomed this as a step forward for ratifying the protection of informal workers. Third, the press, Kompas, Tempo, and VOA news, contain a positive narrative by telling the story of domestic workers and the urgency of their legal protection. Media coverage and narrative are very much in favor of human rights.

Figure 6. Word cloud

From the data processing, it was found that there were 99753 words. The word cloud visualization above shows that words such as "bill", "pprt", "protect", "work", "law", "proposal", "house", and "ladder" are very dominant. This shows that the media narrative highlights the importance of regulation and protection and the mastery of the PPRT Bill. The word protection also indicates that the protection or legal guarantee for domestic workers is an important thing that must be ratified immediately. These dominant words indicate the structural support of the news discourse on the ratification of the PPRT Bill.

In addition to its descriptive function, the TF-IDF weighted word cloud also provides a computational salience map that reflects how discourse attention is mathematically concentrated on specific policy dimensions. By suppressing common but low-informational terms and amplifying issue-specific lexicons, this approach enables scalable identification of dominant narratives in large-scale media corpora, demonstrating the applicability of the model beyond manual qualitative coding.

The main words that often appear are "bill", "pprt", "protect", "work", "law", "proposal", "house", and "ladder". The words "work", "house", and "ladder" form the phrase "domestic worker", which is the main subject of the discourse of the news of the PPRT Bill. The word "pprt," which stands for the protection of domestic workers and is also a word that is often used, strengthens the focus of the main issues of media reporting. The words "bill" and "law" describe how the legal and legislative context is often discussed. And the word "protect" refers to the bill's primary purpose: providing protection and legal guarantees to domestic workers.

In the socio-political context, the words "leadership" and "plenary" appear. House of Representatives, initiatives, debates, and laws. These words represent the legislative process in parliament (DPR), starting from this initiative to the ratification dynamics over the past 2 decades. These words also show that this issue is in the Panjang debate, to show how political correctness affects the ratification of the PPRT Bill. The process is initiated by the House of Representatives to discuss the PPRT Bill at the plenary meeting, and the goal is the law's ratification. However, the socio-political consequences in the last 20 years have affected the stagnation of discussions and even the ratification of the PPRT law. Another word in the socio-political context is the president, Jokowi, Ministers, laws, orders, and the state, which point to the executive, namely the government, in discussing this bill. The role of the president during his answer period, namely Jokowi and the law ministry, was also in the spotlight. The executive and legislative dimensions are the basis for whether they support or slow down the bill's passage.

Another aspect that can be seen is social and gender issues seen from the words "women", "sexual", "violence", "support", "rights", and "human" that appear. The word gives an overview that emphasizes that most domestic workers are women and are vulnerable to violence, as evidenced by the many cases experienced by domestic workers and appearing in the media, as well as reports from the National Commission on Women and data from domestic workers. Human rights are a moral and legal argument to support this bill. The insistence on the ratification of the PPRT Bill also aims to protect domestic workers, most of whom are women, from violence in the domestic realm, both sexual, economic, social, political, physical, and so on.

From a computational perspective, the persistence of gender- and rights-related lexicons across heterogeneous media outlets indicates that the discourse salience patterns are structurally reproduced rather than outlet-specific, suggesting that the model generalizes well across different editorial environments.

Political issues are also a debate, the words that appear a lot are "party", "politics", "camp", "agree", "encourage", "agree", "action". These words can indicate a political tug-of-war within the party and the faction. The action word shows how the Civil Society Movement has supported this bill's immediate passage. This bill is closely related to legal protection, but political dynamics also have an effect, and activism is part of that dynamic.

The words "sigh", "hard", "reaction", "voice", "push", and "exciting" indicate how the public's emotions, as well as the domestic workers themselves and the social movement of the community, began to become frustrated with the slow process of ratification and class support to be passed immediately. The word cloud above shows that the passage of the PPRT Bill is a multidimensional issue that touches on legal, gender, political, and human rights issues. Political dynamics (president, DPR, party, etc.) affect this bill's speed or slower direction. The narrative obtained from the three online media coverage also shows a strong social narrative emerging from women's groups, human rights activists, and public pressure for the House of Representatives to pass the PPRT Bill immediately. Word cloud and sentiment analysis show that the PPRT Bill news is strongly framed by a narrative about legislation or bills and the protection of domestic workers. The existing dynamics show a common thread between politics, gender issues, civil society activities, and struggles.

A more comprehensive analysis shows that, in addition to political aspects or dimensions, gender, human rights, the role of civil society, advocacy actors, and economic factors, there are dimensions of the effect of the media setting and framing agenda. The function of the mass media setting the agenda is the ability of the mass media to select and emphasize several topics, causing the public to accept these topics as necessary [26]. Then, framing analysis is used to see how the media constructs reality and the formation of messages or meanings from the text presented to the audience [27]. This study specifically limited data to online news articles from three mass media, namely Kompas.com, Tempo.com and VOA News Indonesia. This is because these three media are known to have a wide reach and a high history of perspective and credibility. Kompas and Tempo are media that more often contain public policy discourse and VOA News as an international media that has channels in Indonesia and provides reliable viewpoints. Then various ideologies and framing patterns, such as Kompas with moderate, Tempo with critical, and VOA with balanced or global. This diversity is important to map the dynamics of the sentiment of the complex PPRT Bill. In addition, these three media outlets have an online news archive that can be accessed effectively covering the time span needed in this study. Even though the data obtained only exists from 2013 to 2024 because it is known that there is no online version on the three media.

From the news distribution in the data, Kompas and Tempo media report more and tend to be positive. Meanwhile, VOA is more balanced, and this can affect public opinion. The setting of the agenda can be seen from the news, which emphasizes the urgency of protecting and passing the PPRT Bill, thereby encouraging public support. Actors who are pro-abortion against the ratification of the PPRT Bill in their narratives also make their news materials, such as support for figures, research, victim presentations, and data, as a source of public support. Community movements and policy advocacy also signal the existence of a public campaign. Women's organizations and NGOs are the main drivers of the push for the ratification of the PPRT Bill. The coalition of actors, including the media, is a capital to increase the effectiveness of cross-sectoral advocacy. The nature of mass media, which covers society at large and its proximity to people's lives, has a significant influence and makes it possible to carry out the function of political education [28]. This shows that the presence of three online media with its news can influence public opinion.

The dynamics of the ratification of the PPRT Bill is a multidimensional issue. Where there are problems of social and class inequality, gender relations, political battles, human rights struggles, economic injustice and also how the media and civil society have a big role. This PPRT Bill is not only a regulation that is encouraged to be passed immediately, but with the encouragement of this ratification it becomes a symbol of how the struggle for recognition of domestic work is being waged. In addition, it is also protection for vulnerable groups and the renewal of power relations in the personal and national realms. Word cloud from the results of the news above shows that the public narrative is moving towards stronger support. However, political realization and its ratification depend on legislators and power dynamics. The analysis of the above sentiment and the dynamics of news coverage of 387 articles of the PPRT Bill during the period 2004–2024 has successfully implemented and answered the research objectives described. First, the findings of sentiment polarity showed the dominance of aggregate positive sentiment (58.1%), indicating a strong push from the three national online media to support the ratification of the PPRT Bill. Second, the analysis of media coverage patterns identified Tempo as the most active media, confirming that these three major media acted as strong agenda-setting agents in this policy issue. Third, the mapping of sentiment trends chronologically managed to highlight a direct correlation between the peak of negative sentiment spikes in 2019 and positive in 2023 with critical moments in the legislation process. These overall findings not only gauge sentiment, but also confirm the effectiveness of Text Mining in revealing the complex dynamics between the media and public policy, leading to the implication that mass media support is crucial for efforts to accelerate domestic worker protection in Indonesia.

The dominance of positive sentiment (58.1%) in the overall PPRT Bill news coverage indicates strong media support for the ratification of this bill. This aligns with the urgency of PRT protection that has long been voiced by various civil society elements. The peak of negative sentiment in 2019 can be interpreted as a reflection of public and activist disappointment with the slow legislative progress of the PPRT Bill during that period, especially considering the political momentum after the general election. Conversely, the surge in positive sentiment in 2023 is likely related to the designation of the PPRT Bill as a House of Representatives' initiative bill, which offered new hope for its accelerated approval. This dynamic demonstrates that media actively responds to and shapes narratives surrounding the political developments of the PPRT Bill."

Differences in sentiment distribution across media also provide important insights. Kompas, with the highest proportion of positive sentiment (75.9%), appears to take a more proactive role in supporting the RUU PPRT. Meanwhile, VOA News shows a balance of positive and negative sentiment, which may reflect an effort to present a more balanced perspective or report various viewpoints from pro and contra parties. Tempo, as the most active media, shows dominant positive sentiment but with a significant negative portion, indicating broader and possibly more critical coverage of various aspects of the bill's discussion. These findings underscore the media's role as an agent of advocacy and opinion-shaper, where each media outlet may have different editorial tendencies in framing socio-political issues. This cross-media consistency highlights the scalability of the proposed TF-IDF–lexicon framework, which can be directly extended to larger national or cross-country corpora without altering its mathematical structure.

Theoretically, this study enriches the literature on the application of Text Mining and sentiment analysis in the context of protracted legislative issues in developing countries. Practically, these findings can serve as input for policymakers, PRT activists, and civil society organizations to understand how the RUU PPRT is framed in the media, identify critical periods for advocacy intervention, and design more effective communication strategies. Understanding media sentiment can also help mitigate negative narratives and strengthen public support for the PPRT Bill.

Overall, the integration of TF-IDF weighting and lexicon-based sentiment scoring constitutes a reproducible applied-computational framework for longitudinal policy discourse analysis. Beyond the PPRT Bill case, this framework can be utilized to map discourse salience, sentiment concentration, and narrative shifts in other prolonged public-policy debates.

Although this research provides a comprehensive understanding, it is important to acknowledge several limitations. First, lexicon-based sentiment analysis, despite being adapted, has inherent limitations in capturing complex sentiment nuances, sarcasm, or specific cultural contexts in Indonesian. Lexicons may not always cover all sentiment expressions that arise or may misclassify words with multiple meanings. Second, the data corpus is limited to three online media outlets and may not fully represent the entire spectrum of media coverage in Indonesia. Potential biases in media selection or the reporting focus of each media can influence overall sentiment results. Third, this study does not explicitly analyze in-depth external factors that might influence sentiment, such as social media campaigns or politicians' statements, which could be a direction for future research.

4. CONCLUSION

This research successfully analyzed the sentiment and narrative dynamics of PPRT Bill news coverage in three prominent online media outlets in Indonesia over two decades (2004-2024) using Text Mining and Lexicon-Based Sentiment Analysis. The main findings indicate that positive sentiment dominates the news coverage (58.1%), signifying strong media support for this bill. Significant sentiment fluctuations were observed along with political developments, with a peak in negative sentiment in 2019 and a surge in positive sentiment in 2023, reflecting media responses to legislative stagnation and progress. Differences in sentiment patterns across media also highlight the unique role of each outlet in framing this issue. Methodologically, this study contributes a reproducible computational framework that integrates TF-IDF weighting with a discourse-specific sentiment lexicon to enable large-scale, longitudinal sentiment mapping of public policy discourse in the Indonesian language context.

The methodological contribution of this study lies in the adaptation and application of Lexicon-Based Sentiment Analysis tailored for the Indonesian language context in analyzing complex and long-term socio-political issues. This framework allows automated, scalable, and consistent quantification of narrative dynamics across heterogeneous media sources over extended time periods, enabling the identification of diachronic sentiment patterns that are difficult to capture with manual methods. However, this research also has limitations, including the inherent challenges in capturing language sentiment nuances and the corpus's restriction to specific online media.

For future research, it is recommended to expand the data corpus by including more media sources, including social media, and to integrate machine learning-based sentiment analysis methods for validation and comparison. Topic modeling could also be performed to identify more specific narrative themes over time. Thus, a more holistic understanding of the PPRT Bill public discourse can be achieved.

ACKNOWLEDGEMENT

This research was funded by the Vice Chancellor for Research, Innovation, and Entrepreneurship of Satya Wacana Christian University through the Directorate of Research and Community Service for the fundamental research scheme for the 2025 Fiscal Year, Number: 040/SPK-PF/RIK/04/2025.

5. REFERENCES

[1] C. P. F. Anugrah and A. S. Ruslie, "Urgensi Pengaturan dan Perlindungan Hukum Bagi Pekerja Rumah Tangga di Indonesia," *Media Hukum Indonesia*, vol. 2, no. 4, pp. 895–902, Dec. 2024, doi: 10.5281/zenodo.14349190.

[2] T. Kurnianingrum and R. Yamin, "Urgensi Perlindungan Terhadap Pekerja Rumah Tangga," *Info Singkat: Kajian Singkat Terhadap Isu Aktual dan Strategis*, vol. 16, no. 18, pp. 21–25, Sep. 2024.

[3] D. Harahap, "Kekerasan Terhadap PRT Terus Meningkat, Pengesahan RUU PPRT Diminta jangan Gagal lagi," *MediaIndonesia*, Feb. 2024.

[4] C. Putra, F. Anugrah, and A. S. Ruslie, "Media Hukum Indonesia (MHI) Published by Yayasan Daarul Huda Krueng Mane Urgensi Pengaturan dan Perlindungan Hukum Bagi Pekerja Rumah Tangga di Indonesia," vol. 2, no. 4, p. 895, 2024, doi: 10.5281/zenodo.14349190.

[5] D. Putri and V. P. Hanjani, "Ketika Hak Pekerja Rumah Tangga Terabaikan: Potret Realitas Dan Jalan Legislasi Ruu Pprt (Rancangan Undang-Undang Perlindungan Pekerja Rumah Tangga) Yang Mangkrak," *RINEKA: Jurnal Antropologi*, vol. 1, no. 1, pp. 63–74, Jun. 2025.

[6] A. Parvez, A. V. Superani, and I. N. Juaningsih, "Rekonstruksi RUU PPRT Sebagai Upaya Perlindungan Hukum Dalam Penanggulangan Kekerasan Terhadap PRT Perempuan dan Anak," *Ikatan Penulis Mahasiswa Hukum Indonesia Law Journal*, vol. 2, no. 2, pp. 232–250, Feb. 2022, doi: 10.15294/ipmhi.v2i2.54782.

[7] D. N. Kristiyani, "Analisis framing pemberitaan media: dinamika pengesahan RUU PPRT menggunakan model Robert M. Entman," *Jurnal Komunikasi Universitas Garut: Hasil Pemikiran dan Penelitian*, vol. 11, no. 2, pp. 603–621, Oct. 2025.

[8] P. Laksono, I. K. Pesantren Abdul Chalim, and K. Massa, "Kuasa Media dalam Komunikasi Massa," vol. 4, no. 2, pp. 49–61, Oct. 2019.

[9] Sri Choiriyati, "Peran Media Massa dalam Membentuk Opini Publik," *Perspektif*, vol. 2, no. 2, 2015.

[10] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, "Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization," in *2020 5th International Conference on Informatics and Computing, ICIC 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICIC50835.2020.9288655.

[11] F. Aftab, S. U. Bazai, L. Baloch, S. Aslam, Amphawan Angela;, and T. K. Neo, "A Comprehensive Survey on Sentiment Analysis Techniques," *International Journal of Technolgy (IJTech)*, vol. 14, no. 6, pp. 1288–1298, Oct. 2023, doi: https://doi.org/10.14716/ijtech.v14i6.6632.

[12] D. Tribuana, U. Usman, and D. Dayanti, "Penerapan Natural Language Processing Untuk Analisis Sentimen Terhadap Layanan Publik Di Media Sosial Twitter," *Jurnal Teknologi dan Bisnis Cerdas*, vol. 1, no. 1, pp. 28–37, Jul. 2025, doi: 10.64476/jtbc.v1i1.3.

[13] S. Vijay Gaikwad, A. Chaugule, and P. Patil, "Text Mining Methods and Techniques," *Int J Comput Appl*, vol. 85, no. 17, pp. 975–8887, Jan. 2014.

[14] Y. A. Hafiz and E. Sudarmilah, "Impelementasi Web Scaping Pada Portal Berita Online," *Inisiasi: Jurnal Inovasi dan teknologi*, vol. 12, Oct. 2023.

[15] A. Ahmed and A. Salam, "Automatic Scientific Literature Gathering and Analysis from Textual Corpus using Web Scraping and Locality Sensitive Hashing," *IEEE-SEM*, vol. 11, no. 8, pp. 75–80, Aug. 2023.

[16] M. Aqib, R. Mehmood, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, "Smarter traffic prediction using big data, in-memory computing, deep learning and gpus," *Sensors (Switzerland)*, vol. 19, no. 9, May 2019, doi: 10.3390/s19092206.

[17] H. Suyal, A. Panwar, and A. S. Negi, "Text Clustering Algorithms: A Review," *International Jurnal of Computer Applications*, vol. 96, no. 24, 2014.

[18] S. R. Shah, A. Kaushik, S. Sharma, and J. Shah, "Opinion-Mining on Marglish and Devanagari Comments of YouTube Cookery Channels Using Parametric and Non-Parametric Learning Models," *big data and cognitive comouting*, vol. 4, no. 3, pp. 1–19, Mar. 2020, doi: https://doi.org/10.3390/bdcc4010003.

[19] T. Bintang, S. Silalahi, and A. Toni, "Pengaruh Pro & Kontra Pilkada 2020 Pada Media Sosial Twitter (Drone Emprit: Pilkada 2020-Pro & Kontra)," *Communication*, vol. 12, no. 2, pp. 143–153, Oct. 2021.

[20] M. K. A. Reiki, Y. Sibaroni, and E. B. Setiawan, "Comparison of Term Weighting Methods in Sentiment Analysis of the New State Capital of Indonesia with the SVM Method," *International Journal on Information and Communication Technology (IJoICT)*, vol. 8, no. 2, pp. 53–65, Jan. 2023, doi: 10.21108/ijoict.v8i2.681.

[21] M. Kayest and S. K. Jain, "Optimization driven cluster based indexing and matching for the document retrieval," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 3, pp. 851–861, Mar. 2022, doi: 10.1016/j.jksuci.2019.02.012.

[22] L. C. Chen, "An extended TF-IDF method for improving keyword extraction in traditional corpus-based research: An example of a climate change corpus," *Data Knowl Eng*, vol. 153, Sep. 2024, doi: 10.1016/j.datak.2024.102322.

[23] I. Arroyo-Fernández, C. F. Méndez-Cruz, G. Sierra, J. M. Torres-Moreno, and G. Sidorov, "Unsupervised sentence representations as word information series: Revisiting TF–IDF," *Comput Speech Lang*, vol. 56, pp. 107–129, Jul. 2019, doi: 10.1016/j.csl.2019.01.005.
[24] L. Gomes, R. da Silva Torres, and M. L. Côrtes, "BERT- and TF-IDF-based feature extraction for long-lived bug prediction in FLOSS: A comparative study," *Inf Softw Technol*, vol. 160, pp. 1–2, Aug. 2023, doi: 10.1016/j.infsof.2023.107217.
[25] H. Abkoriyah and T. T. Dewi, "Objektivitas Berita di Harian Kompas dan Kompas.com," *CoverAge: Journal of Strategic Communication*, vol. 7, no. 2, 2017.
[26] J. P. Tambusai, E. Efendi, A. Taufiqurrohman, T. Supriadi, E. Kuswananda, and K. P. Islam, "Teori Agenda Setting," *Jurnal Pendidikan Tambusai*, vol. 7, no. 1, 2023.
[27] E. L. Pambayun, *One Stop Qualitative Research Methodology In Communication*. Jakarta: Lentera Ilmu Cendekia, 2018.
[28] A. Aryanti and H. Rusitawati, "Netralitas Media Massa Sebagai Impelemntasi Fungsi Edukasi Politik di Indonesia," *Jurnal Penelitian Politik*, vol. 2, 2014.