



Text Clustering as A Computational Approach for Analyzing the Framing of Election News on Detik.com

Atik Ariesta 

Budi Luhur University, South Jakarta 12260, Indonesia
atik.ariesta@budiluhur.ac.id

Abstract. This research develops a computational approach based on text clustering to analyze election news framing on the Detik.com portal. In the context of elections, media framing becomes an essential instrument in shaping public opinion towards political issues and actors. This research applies the K-Means algorithm to group news based on the similarity of text patterns in news titles, using TF-IDF weighting for feature representation. Data were collected through web scraping from the news.detik.com/pemilu channel between June 2023 and January 2024, resulting in more than 16,000 news items for analysis. This research also includes the development of a web application prototype using the Flask framework, which automates the text clustering process and supports framing analysis. This application offers features for exploring clustering results and word cloud visualization, enabling researchers to identify framing elements based on Entman's theory—specifically, defining problems, diagnosing causes, making moral judgments, and recommending treatments. Cluster quality evaluation was conducted using the Davies-Bouldin Index (DBI), which yielded consistent cluster results and warranted further analysis. The results indicate that the clusters formed represent various news patterns, such as candidate support narratives, coalition dynamics, and election regulations. This research not only offers a quantitative approach to media framing analysis but also produces an intelligent system-based tool that can be replicated in other political text studies. The findings broaden the understanding of the application of text mining techniques for media framing analysis in the realm of digital political communication.

Keywords: text clustering, media framing, election, k-means, web application

1 Introduction

The media plays a vital role in shaping public understanding of political issues, especially during elections. Framing in news reporting is a technique that highlights certain aspects of reality while minimizing others to influence how the public views an issue. Entman (1993) Framing theory identifies four main functions in media discourse: defining problems, diagnosing causes, making moral judgments, and proposing solutions.

In the context of reporting on the 2024 Indonesian general election, Detik.com, a leading online news outlet, has become a primary source of political information for the public. The media plays a vital role in shaping public perceptions of candidates,

political issues, and election results. As one of the most influential news sites in Indonesia, Detik.com significantly influences public opinion. According to data from Similarweb, Detik.com ranks second in the News & Media Publisher category for website visits.

Although online news media have a significant influence on political perceptions, research on elections in Indonesia has focused mainly on qualitative framing analysis or text classification methods such as sentiment analysis. Few studies have integrated computational unsupervised learning classification methods with framing theory to identify framing in election headlines. This gap is evident in the context of Indonesian-language media and large-scale online news outlets.

This research used an unsupervised learning approach, specifically the K-Means algorithm, because this algorithm is one of the most popular and easily understood clustering methods (Hasan et al., 2022; Ikotun et al., 2023). Previous research has shown that K-Means is effective in clustering Bangla-language news documents using the TF-IDF approach. Simanjuntak et al. (2023) Also applied K-Means to analyse clusters of text data from social media, specifically Twitter. Their results showed dominant clusters related to economics, politics, social issues, and sports. The analysis was conducted using Orange Data Mining, including lexicon-based sentiment analysis. Disayiram & Rupasingha (2022) examined the application of clustering algorithms to group English-language news articles into five main categories: politics, sports, health, technology, and business. News articles were converted into vectors using the TF-IDF method, then clustered using three algorithms: Expectation-Maximization (EM), Simple K-Means, and Hierarchical Clustering. Evaluation using WEKA showed that the EM algorithm provided the highest accuracy of 88.5%. This research also concluded that news content provided more accurate clustering results than using news headlines alone. Hamami & Dahlan (2024) developed an online news segmentation system based on the K-Means algorithm combined with the Elbow method. This system groups news based on content similarity from the Detik news portal. The number of clusters changes daily following emerging topic trends, and the segmentation results are visualized through a web application, making it easier for users to access popular topics without having to read the entire news story.

Although previous research has demonstrated the effectiveness of clustering algorithms such as K-Means for topic detection and news segmentation, most studies have emphasized thematic grouping rather than media interpretation. This results in clustering results rarely being linked to framing construction. Integrating classification with Entman's framing theory enables the identification of framing structures beyond topic clustering.

Ashfaq et al. (2022) examined the use of framing in Pakistani media coverage of political protests spread on Twitter. They collected 498 tweets from three major media outlets (Dawn, The News, and Express Tribune) between October 20 and November 20, 2019. They conducted a quantitative framing analysis based on thematic and episodic categories, tweet types, and subject matter. Meanwhile, a study by Anggoro et al. (2023) analyzed media framing of women's representation as Indonesian presidential candidates in 2024 using Robert Entman's model and a qualitative approach. The analysis was conducted on articles from Republika.com and Sindonews.com. The results

show that Sindonews.com highlights the presence of female candidates positively, but is still limited to gender stereotypes as vice presidential candidates. At the same time, Republika.com places greater emphasis on electability and objective party support. Both media outlets contribute to shaping public perception of female candidates, despite bias in their reporting.

Existing framing research has largely relied on manual analysis and qualitative interpretation, yielding rich contextual insights but limiting scalability and efficiency. In contrast, computational approaches allow for large-scale processing of news data. However, research aligning clustering results with framing theory for Indonesian online news media remains rare.

In general, framing analysis is still primarily conducted manually, which is time-consuming and tends to be subjective. Therefore, computational approaches such as text clustering are needed to provide a more efficient and objective analysis alternative. Furthermore, the development of a web-based application prototype that can automate news processing and facilitate systematic framing analysis is also needed.

Therefore, this study aims to apply the K-Means Clustering algorithm to group Detik.com news headlines related to the 2024 Indonesian elections, use the Clustering results to identify framing using Entman's framing theory, and develop a web-based prototype to support the Clustering process and framing analysis. The contribution of this study is to operationalize framing theory within a computational framework and to apply unsupervised learning and text mining to political communication analysis.

2 Methodologies

2.1 Research Framework

This research adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which consists of: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Implementation (Wirth & Hipp, 2000).

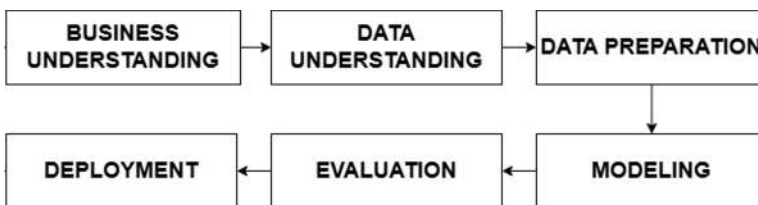


Fig. 1. CRISP-DM Steps

The first step is business understanding. In this step, the researcher determines the research objectives and identifies the problem to be solved. In this research, the problem identification is the clustering of election news for the analysis of election news framing on the election news website, news.detik.com/pemilu. Framing is conducted to under-

stand how the mass media actively selects and emphasizes certain aspects to shape public perception (Yazmi et al., 2024). This step will result in the determination of the main research problem related to the clustering of election news texts for framing analysis.

The second step is data understanding. Election news was collected from news.detik.com/pemilu using web scraping. The procedure for automatic extraction of data from websites using software (Khder, 2021). The tool used is a web scraper, an add-on for Google Chrome. The data to be collected covers news from June 2023 to January 2024. This timeframe was chosen because this period is the lead-up to the presidential and vice-presidential elections, which are characterized by escalating political issues, campaign strategies, and high media attention to the democratic process.

The third step is data preparation. Researchers conduct text processing, consisting of tokenization, stop word removal, and stemming (Jo, 2019). The text to be processed is a news headline. Tokenization involves removing special characters and numbers, converting to lowercase, and breaking the text based on whitespace. Stop word removal involves removing irrelevant words that frequently appear in the text, such as prepositions and conjunctions. Stemming is performed based on the results of stop word removal. Stemming converts each token containing an affix into its root form. The stemming process uses Sastrawi, the most appropriate stemmer for Indonesian (Firman Sodik et al., 2022). After data processing, researchers weigh words using Term Frequency-Inverse Document Frequency (TF-IDF). The output of the third step is a news headline that has undergone text processing and a TF-IDF weight matrix for each word in the headline. After the processing stage, the dataset is checked for duplicate titles and incomplete notes. The inspection results indicated that no duplicate data was identified, and all collected news titles met the criteria to move to the next stage. As a result, the total number of news titles remains unchanged after pre-processing, totaling 16,259 titles from June 2023 to January 2024.

The fourth step is modelling. Researchers cluster the headlines, which have already received their weighted values. Before carrying out the clustering process, the optimal k value is determined using the elbow method. The elbow shape of the elbow method graph is the k value that will be used to determine the cluster value. The output of this step is the optimal number of clusters determined using the elbow method, and news title clustering is carried out based on the TF-IDF value. The K-Means algorithm works by randomly initializing k cluster centers and then allocating each data item to a cluster based on its proximity to the nearest cluster center (Widaningrum et al., 2022).

Although K-Means is more efficient and widely used, it has several limitations, including sensitivity to the initial selection of centroids, which can cause clustering results to vary over the course of the process. Additionally, K-means assumes spherical cluster shapes and equal cluster sizes, so it does not fully capture the semantic relationships in political news headlines. Research by Khan et al. (2024) emphasize that K-Means is strongly influenced by outliers and requires determining the number of clusters (K), which affects the quality of clustering and the level of interpretation. Furthermore, research by Li et al. (2024) and Semoglou et al. (2025) stated that the K-Means algorithm is less suitable for high-dimensional text data without determining appropriate parameters. Therefore, the existing limitations are taken into consideration when interpreting the results and framing the analysis.

The fifth step is evaluation. Clustering results are evaluated using the Davies-Bouldin Index (DBI) to measure the quality and suitability of the clusters generated by the K-Means algorithm. The DBI for a given cluster is calculated as the ratio of Intra-Cluster Dispersion to Inter-Cluster Distance using Euclidean Distance (Anita & Aliando, 2022). The result of this step is the DBI value for each dataset. DBI offers an objective and quantitative way to evaluate clustering by measuring the similarity and separation of clusters. This approach is suitable for the K-Means Clustering method (Chicco et al., 2025). DBI has been demonstrated as a useful tool for assessing clustering performance in text data by identifying clusters that are internally scattered and close to other clusters (Ikotun et al., 2025). According to Hassan et al. (2024), the clustering validity index, including DBI, is a key tool for determining the optimal number of clusters and comparing classification results. Using DBI for evaluation ensures that the chosen clustering method has good internal cohesion and separation, which supports interpretation in text-based analysis.

The sixth step is deployment. In this step, researchers select the cluster with the most significant amount of data each month. From these selected clusters, researchers identify the framing dimensions and framing elements. The framing dimensions include issue selection and aspect emphasis. The framing elements include defining the problem, diagnosing the cause, making moral judgments, and recommending treatment. Defining the problem involves clarifying the issue, showing how a research issue can be viewed, and identifying the nature of the problem. Diagnosing the cause involves estimating the source of the problem, understanding the events caused by other events, and pinpointing the actors responsible. Making moral judgments involves evaluating the moral values used to legitimize or delegitimize actions. Treatment recommendations consist of proposed solutions and pathways to address the issue (Boer et al., 2020). The outcome of this step is the framing dimensions and framing elements in election news. Alignment between cluster themes and Entman's framing theory is conducted to enhance the validity of framing analysis results and minimize subjective bias in the qualitative interpretation of cluster findings.

2.2 Prototype

A web-based prototype was developed using the Flask framework. It features tools for text processing, text clustering, and framing data analysis. The prototype serves as a platform for exploring framing patterns in political news. The prototype integrates several functional modules, including media input data management, text preprocessing (tokenization, stop word removal, and stemming), TF-IDF vectorization, K-Means clustering, and framing analysis visualization. The system architecture allows users to upload news data, run a dynamic clustering process, and examine dominant clusters and their corresponding framing elements. This prototype is designed to support reproducibility and facilitate exploratory analysis for researchers and media analysts.

3 Result and Discussion

3.1 Data Collection

Data shows a significant upward trend from June 2023 to January 2024, with the initial count of 1,194 cases almost tripling to 3,111. A sharp spike happened from September to December 2023, indicating rapid growth during that period. Overall, the total for the eight months reached 16,259.

Table 1. Number of news stories per month

No	Month Year	Amount
1	Juni 2023	1.194
2	Juli 2023	1.279
3	Agustus 2023	1.613
4	September 2023	1.910
5	Oktober 2023	2.308
6	November 2023	2.185
7	Desember 2023	2.659
8	Januari 2024	3.111
Total		16.259

Table 2 presents the attributes produced from web scraping on the Detik.com website, specifically news.detik.com/pemilu. Eight attributes were collected using the Google Chrome web scraper tool.

Table 2. Attribute Scrapping

No	Attribute	Description
1	<i>web-scraper-order</i>	The sequence of web scraper tools scraping news data
2	<i>web-scraper-start-url</i>	The initial URL link of the web scraper scraping news data
3	News	Contains the date and title of the scraped news story
4	news-href	The URL link of the scraped article/news story
5	Title	The title of the scrapped article
6	Date	The date of the scrapped article
7	NewsPicture-src	The main image link of the scraped news story
8	PictureCaption	The short text of the main image scraped
9	NewsContent	The content of the scraped news story
10	Page	The page of the scrapped news story in the news list
11	<i>Page-href</i>	The URL link of the scraped news story

Table 3 lists the attributes used for clustering: News Date, News Title, News Link, News Image Link, and News Content. These attributes were chosen based on the requirements of the news framing analysis. The clustering focused on the news title at-

tribute because titles often include keywords and the media's perspective, which indicate framing direction. As a result, titles are very useful for identifying the topic and framing elements in the analysis.

Table 3. Attributes Used

No.	Dataset Attribute	Description
1	Date	News Date
2	Title	News Title
3	News Link	News Link
4	Picture Link	News Main Image Link
5	News Content	News Content

Table 4 is the Detik news dataset for June 2023, which represents one of the eight datasets generated.

Table 4. Detik News Research Dataset for June 2023

Date	Title	News Link	Image Link	News Content
6/1/2023	Romahurmuziy: Sandiaga Uno Sudah 96 Persen Gabung PPP	https://news.detik.com/pemilu/d-6749932/romahurmuziy-sandiaga-uno-sudah-96-persen-gabung-ppp	https://akcdn.detik.net.id/community/media/visual/2023/06/01/romahurmuziy_169.jpg?w=700&q=90	Jakarta - The Chairman of the United Development Party (PPP) Advisory Council, Romahurmuziy, stated that Sandiaga Uno is just waiting for the right moment to join the PPP. Romy also stated that Sandi is 96 percent certain to join the PPP.....
.
.
.
6/30/2023	Beredar SE Walkot Depok Soal Penertiban Spanduk hingga Atribut Partai	https://news.detik.com/pemilu/d-6799957/beredar-se-walkot-depok-soal-penertiban-spanduk-hingga-atribut-partai	https://akcdn.detik.net.id/community/media/visual/2023/06/23/walkot-depok-m-idris-devi-detikcom_169.jpg?w=700&q=90	Jakarta - The Chairman of the United Development Party (PPP) Advisory Council, Romahurmuziy, stated that Sandiaga Uno is just waiting for the right moment to join the PPP. Romy also stated that Sandi is 96 percent certain to join the PPP.....

3.2 Text Processing

Text processing started with tokenization, stop word removal, and stemming. The attribute used in the study was the title. Table 5 shows an example of removing special characters and numbers, converting to lowercase, and tokenization (Jo, 2019).

Table 5 shows the text processing of the initial title, which is then transformed by removing special characters and numbers and converting it to lowercase. The Transform Text column displays the removal of special characters such as double quotation marks (") and colons (:), and the text has been converted to all lowercase. The Tokenization column shows the results of splitting the news titles into tokens based on whitespace.

Table 5. Transform and Tokenize News Titles

Initial Title	Transform Text	Tokenization
Criticizing Jokowi, Puskapol UI: The Excuse of "cawecawe" Pilpres is for "Nation and State" a Cliché Excuse	criticizing jokowi puskapol ui excuse cawecawe pilpres for nation and state cliché excuse	['criticizing', 'jokowi', 'puskapol', 'ui', 'excuse', 'cawecawe', 'pilpres', 'for', 'nation', 'and', 'state', 'cliché', 'excuse']

Table 6 shows the results of tokenization, which will undergo stopword removal. As seen in the Stopword Removal column, the words "untuk," "cawecawe," and "dan," which are conjunctions, are no longer present. Stemming also merges tokens into a single sentence.

Table 6. Stopword Removal News Title

Tokenization	Stopword Removal
['criticizing', 'jokowi', 'puskapol', 'ui', 'excuse', 'cawecawe', 'pilpres', 'for', 'nation', 'and', 'state', 'cliché', 'excuse']	['criticizing', 'jokowi', 'puskapol', 'ui', 'excuse', 'pilpres', 'nation', 'state', 'cliché', 'excuse']

Table 7. Stemming News Title

Stopword Removal	Stemming
['criticizing', 'jokowi', 'puskapol', 'ui', 'excuse', 'pilpres', 'nation', 'state', 'cliché', 'excuse']	criticizing jokowi puskapol ui excuse pilpres nation state cliché excuse

Table 7 displays the stemming results after stop word removal and combining each token into a sentence. Following stemming, the word "alasan" is replaced with its root word, "alas," as shown in the stemming column.

The next step in text processing is to weight the words in the news titles. Table 8 shows an example of TF-IDF results from three news items: News 1, News 2, and News 3. Cells with a number indicate the news item contains the word corresponding to the

column, while cells with a zero indicate the news item does not contain the word corresponding to the column. News 1 contains the words "ada", "anies", "democrat", "shock", "pacitan", and "sby". News 2 contains the words "electability", "ganjar", "litbang", "win", "prabowo", "overtake", "relax", "moment" and "survey". News 3 has the words "Airlangga", "open", "caleg", "want", "claim", "PDIP", "election", "propotional", "system".

3.3 Clustering

The clustering process starts by determining the number of clusters. Then, it proceeds with the clustering using the K-Means algorithm. Finally, the evaluation is done with the DBI algorithm.

Table 8. Term Frequency – Inverse Document Frequency (TF-IDF) News Title

Words	News 1	News 2	News 3
ada	0,4041	0	0
airlangga	0	0	0,3842
anies	0,2317	0	0
caleg	0	0	0,3438
claim	0	0	0,3484
democrat	0,2427	0	0
electability	0	0,2733	0
election	0	0	0,2328
ganjar	0	0,1448	0
litbang	0	0,3609	0
meet	0,2661	0	0
moment	0	0,3783	0
open	0	0	0,3394
overtake	0	0,4028	0
pacitan	0,4774	0	0
pdip	0	0	0,1956
prabowo	0	0,2171	0
propotional	0	0	0,3586
relax	0	0,3783	0
sby	0,357	0	0
shock	0,4774	0	0
suestion	0,2647	0	0
survei	0	0,2809	0
system	0	0	0,3277
want	0	0	0,4114
win	0	0,2634	0

Figure 2 displays eight elbow method graphs from Datasets 9 to 16 for Detik News. The first graph (top left) shows the elbow for the optimal cluster in June with a value

of $k = 12$. The second graph (top right) shows the elbow for the optimal cluster in July with a value of $k = 24$. The third graph (second row, left) displays the elbow for August with $k = 12$. The fourth graph (second row, right) depicts the elbow for September with $k = 24$. The fifth graph (third row, left) illustrates the elbow for October with $k = 20$. The sixth graph (third row, right) shows the elbow for November with $k = 14$. The seventh graph (fourth row, left) presents the elbow for December with $k = 13$. The eighth graph (fourth row, right) displays the elbow for January with $k = 15$.

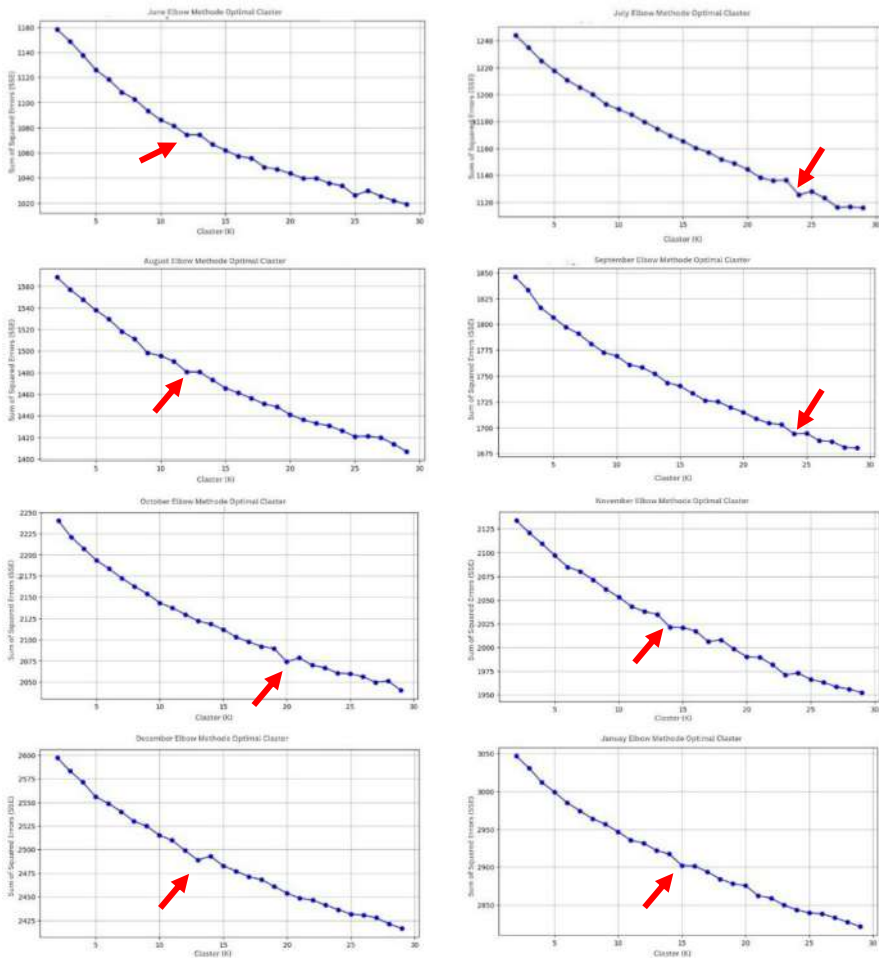


Fig. 2. Elbow Method Chart Detik News

data size for each cluster from the Detik News Dataset. The first graph (top-left) shows the data for each cluster in June, from cluster 0 to 11 (12 clusters total). The second graph (top-right) depicts the data in July, from cluster 0 to 23 (24 clusters). The third graph (bottom-left) presents the data for August, from cluster 0 to 11 (12 clusters). The fourth graph (bottom-right) illustrates the data for September, from cluster 0 to 24 (25 clusters). The fifth graph (third row, left) shows the data for October, from cluster 0 to 189 (20 clusters). The sixth graph (third row, right) displays the total data for November, from clusters 0 to 13 (14 clusters). The seventh graph (fourth row, left) shows the total data for December, from clusters 0 to 12 (13 clusters). The eighth graph (fourth row, right) depicts the total data in January, from clusters 0 to 14 (15 clusters).

The clustering results were assessed using the Davies-Bouldin Index (DBI) to evaluate the quality and appropriateness of the clusters created by the K-Means algorithm. Table 9 presents the evaluation outcomes for each cluster. The DBI results indicate that each optimal cluster, determined by the elbow method, has the lowest DBI value.

Table 9. Davies-Bouldin Index (DBI) Value Detik News

No.	Dataset	k	Davies-Bouldin Index (DBI)
1	News Detik Juni 2023	12	5,38
2	News Detik Juli 2023	24	5,12
3	News Detik Agustus 2023	12	6,04
4	News Detik September 2023	24	5,29
5	News Detik Oktober 2023	20	5,77
6	News Detik November 2023	14	6,41
7	News Detik Desember 2023	13	6,95
8	News Detik Januari 2024	15	6,72

The DBI value in Table 9 is relatively high, but it was derived from news headline text. News headlines often share similar semantics, recurring political figures, and the same keywords. This leads to increased intra-cluster variance and decreased inter-cluster separation.

3.4 Data Presentation

Data presentation was conducted by selecting the cluster with the largest amount of data each month. These selected clusters were then analyzed for dimensions and framing elements. Figure 3 displays the clusters with the highest number of news items for each month. Table 10 displays the clusters with the highest number of news items from June 2023 to January 2024, based on the Detik news clustering results.

Table 10. Cluster with the Most News

No.	Month	Cluster	News Amount
1	June 2023	4	430
2	July 2023	11	188
3	August 2023	2	558

No.	Month	Cluster	News Amount
4	September 2023	16	427
5	October 2023	2	456
6	November 2023	7	407
7	December 2023	8	594
8	January 2024	0	728

Table 11 presents the results of the initial stage of framing dimensions, specifically, news issue selection. The issue selection process was based on clustering results, which include keyword frequency and topic coherence within each cluster. Prominent issues from each cluster were identified by examining the number of news items and frequently recurring words. To verify the accuracy of the framing dimension interpretation, political communication experts conducted the analysis.

Table 11. Detik Political News Issue Selection


No.	Month	Selecting Issue
1	Juni 2023	<ul style="list-style-type: none"> a. Elections and National Political Dynamics b. Internal Dynamics of Political Parties c. Support from Volunteers and Community Leaders d. Social and Government Issues e. Surveys and Electability
2	Juli 2023	<ul style="list-style-type: none"> a. Coalitions and Political Strategy b. Gender and Representation c. National Policies and Issues d. Viral Phenomena and Public Response e. Regional Head Elections f. Social and Cultural Issues
3	Agustus 2023	<ul style="list-style-type: none"> a. Presidential and Vice-Presidential Candidates b. Coalition Dynamics and Party Relations c. Election Strategy and Campaign d. Internal Controversy and Criticism e. Policies and Amendments f. Regional Issues and Local Figures
4	September 2023	<ul style="list-style-type: none"> a. 2024 Presidential Election Issues b. Identity Issues and Political Polarization c. Maneuvers of Political Parties and Political Figures d. Social Dynamics and Public Support e. Special Issues and Supporters
5	Oktober 2023	<ul style="list-style-type: none"> a. The Political Dynamics of the 2024 Presidential and Vice-Presidential Candidates b. The Battle of Narratives and Campaign Strategies c. External and Internal Party Conflicts d. Regulation and Neutrality in the Election e. Volunteer Visits and Declarations
6	November 2023	<ul style="list-style-type: none"> a. The Dynamics of Jokowi's Relationship with the Indonesian Democratic Party of Struggle (PDIP) b. Kaesang and the PSI's Political Maneuvers c. Election and Candidacy Context d. Internal and Cross-Party Conflict


No.	Month	Selecting Issue
7	December 2023	e. Megawati's Speeches and Statements
		a. Presidential Candidate Campaigns and Programs
		b. Elite Political Relations
		c. Social and Security Issues
		d. Surveys and Electability
8	Januari 2024	e. Regional-Specific Issues
		a. Political Campaigns and Promises
		b. Political Strategy and Elections
		c. Responses to National Issues
		d. Religion and Traditions
		e. Relationships with Other Figures and Parties

After selecting the cluster with the largest number of news items, the researchers analyzed the framing dimensions and elements. The framing dimensions involved choosing issues and emphasizing aspects. The framing elements included defining the problem, diagnosing the cause, making moral judgments, and suggesting treatment options. The analysis of the framing dimensions and elements involved interpretative factors influenced by the researcher's subjectivity. Therefore, this analysis was conducted systematically using Entman's framework, with input from several experts.

Table 11 shows a selection of issues for each month. From June 2023 to January 2024, significant developments in Indonesian politics occurred as the country moved toward the 2024 elections. The focus initially was on election preparations and internal party dynamics, then shifted to coalition building and political strategies after the announcement of presidential and vice-presidential candidates. As the election drew near, intense discussions centered on campaign tactics, narrative battles, and internal and cross-party conflicts. Key political figures and elite relations' dynamics dominated news, including matters related to election regulations and neutrality. Finally, as election day approached, attention moved to campaign promises, final strategies, and responses to national issues, marking a critical phase in political competition. January 2024 wrapped up the series with a focus on election campaigns, public figure involvement, and socio-economic and political controversies.

Table 12. Highlighting Aspects of Detik News

Date, News Title	Highlighting Aspects of Title Image	Emphasis on Word/Sentence Aspects
1 Juni 2023, "Ganjar Calls for United Support from Parties and Volunteers in the 2024 Presidential Election"		"Ganjar emphasized the importance of collaboration between various elements in winning the election. He called on all parties to work together to achieve a shared victory. "Today, we are demonstrating to the public how collaboration, cooperation, and mutual respect are carried out because we share the same goal: victory," he said.

Date, News Title	Highlighting Aspects of Title Image	Emphasis on Word/Sentence Aspects
2 Juni 2023, "The PAN Chairman Will Visit PDIP Headquarters This Afternoon to Discuss Support for Ganjar?"		"PAN has handed over the presidential candidacy issue to Zulhas. "So, PAN will also have a meeting tomorrow night at the Central Executive Board (DPP), one of which will discuss current issues, including the presidential election contest, even though PAN has given its full mandate to the General Chairperson," Yandri said. "However, updates or current developments need to be discussed internally within the party, and we are still communicating with all parties."

After selecting the issues, the second and final stage of the framing dimension is aspect emphasis. Aspect emphasis is analyzed based on the use of specific words, sentences, images, and visuals presented by the media for audience consumption. Table 12 provides a sample of news aspect emphasis drawn from a selection of news issues from each month of Detik's election coverage.

In Table 12, the aspect highlighting is based on an analysis of news items that have been clustered and selected during the issue selection stage. This process involves identifying visual elements and the language used by the media, selecting news headlines, utilizing headline images, and incorporating sentences from the news content.

Table 13. Detik News Framing Elements

News Title	Define Problem	Diagnose Cause	Make a Moral Judgment	Treatment Recommendation
Ganjar Calls for United Support from Parties and Volunteers in the 2024 Presidential Election	Ganjar Pranowo called for collaboration between political parties and volunteers to win the 2024 presidential election. The main problem identified was the potential for disharmony or lack of coordination between political parties and volunteers in supporting presidential candidates, which could reduce campaign	While political parties have formal structures and organized strategies, volunteers are usually independent and more flexible, making it sometimes difficult to align the goals and working methods of the two parties.	Ganjar's call for collaboration between political parties and volunteers is a positive step, demonstrating the spirit of mutual cooperation (gotong royong) democracy. However, it is crucial to ensure that this collaboration is not solely aimed at political victory, but also reflects values that respect people's sovereignty,	Form a liaison team between political parties and volunteers to align campaign strategies. Ensure that all parties understand and consistently support the candidate's vision and mission. Most importantly, provide training to volunteers to effectively convey campaign messages without compromising democratic values.

News Title	Define Problem	Diagnose Cause	Make a Moral Judgment	Treatment Recommendation
	effectiveness and the chances of victory.		openness, and inclusiveness. A balance between party leadership and volunteer aspirations must be maintained to avoid alienating either party.	
PAN Chairman to Visit PDIP Headquarters This Afternoon to Discuss Support for Ganjar?	The meeting between PAN General Chair Zulkifli Hasan and PAN administrators with PDIP revealed efforts to gather support for presidential hopeful Ganjar Pranowo. The primary issue is the uncertainty about PAN's political stance in the 2024 presidential election, which could impact the party's internal unity and its relations with the coalition partners.	The presidential candidate's decision is entirely in the hands of the PAN General Chair, which can create internal tensions if it is not supported by party consensus	PAN's choice to engage in dialogue with PDIP is a positive move toward supporting democratic values through political communication. However, transparency with the public and consistency with coalition partners must be maintained to prevent confusion or the appearance of opportunism. The decision-making process should also involve inclusive party mechanisms to ensure collective agreement and avoid internal disagreements	PAN must choose its presidential candidate within a clear deadline, as promised by the Deputy Chairperson of PAN. This will give political certainty to party members and supporters

Table 13 provides an analysis of various election-related news stories in Indonesia using a framing approach with four dimensions: Define Problem, Diagnose Cause, Make Moral Judgment, and Treatment Recommendation. Each framing dimension is developed by examining news content. The analysis focuses on identifying news texts that address the main problem (Define Problems), explain the cause (Diagnose Cause), offer moral judgments (Moral Judgment), and suggest solutions or actions (Treatment Recommendation).

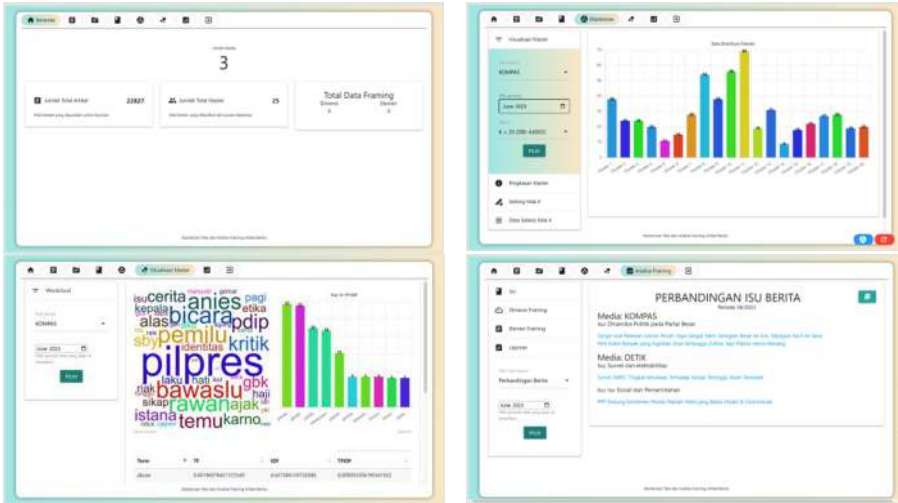


Fig. 4. Application Screenshot

Figure 4 shows a screenshot of the application interface to illustrate the system's functionality. While the main analysis centers on Detik.com headlines, the user interface screenshots in this paper display test data from Kompas.com for layout and visual validation. Compared to manual framing analysis, this computational method offers increased speed and scalability, while remaining interpretable through visualization and clustering validation.

3.5 Discussion

The clustering process using the K-Means algorithm showed successful results, although it still has some limitations. Using TF-IDF together with K-Means offers an efficient way to cluster news articles. The Silhouette Score indicates that the chosen number of clusters is optimal; however, despite the algorithm effectively grouping news headlines, overlaps still occurred. This issue is consistent with K-Means' known limitations, such as its inability to form spherical clusters and handle irregular semantic boundaries effectively.

Parameter setting, especially choosing k , is essential. The Elbow Method and Silhouette analysis were used to select k , but alternative clustering methods like DBSCAN or Hierarchical Clustering might produce different results, particularly for overlapping topics. The text preprocessing steps, including tokenization, stopwords removal, and stemming, played a vital role in preparing the data for clustering. However, preprocessing can unintentionally exclude contextually important words, affecting the outcomes. From a practical standpoint, clustering results provide a useful basis for framing analysis. By grouping articles on election themes such as candidate competition, coalition dynamics, and regulatory issues, the K-Means algorithm helps enable a more struc-

tured interpretation of media framing patterns. Still, the accuracy of framing interpretation relies on qualitative analysis, underscoring the importance of combining computational techniques with domain expertise.

Future improvements might include integrating semantic embeddings like Word2Vec or BERT to better understand contextual relationships, as well as using semi-supervised clustering to include partial labels from framing experts. These enhancements could help balance computational efficiency with interpretative depth, resulting in outcomes that are both technically strong and meaningfully insightful.

4 Conclusion

The K-Means algorithm was effectively used to group election news on Detik based on text patterns, which were then employed for framing analysis. The findings show that the media actively select and emphasize certain aspects to influence public perceptions of political issues. The clustering results reveal a distribution of issues that mirror political dynamics, such as candidate competition, coalition matters, and election regulation policies. The framing elements offer a comprehensive view of how the media define problems, identify causes, make moral judgments, and suggest solutions.

Future research should expand news data collection across more media platforms and over a longer timeframe to achieve more representative results. Additionally, incorporating sentiment analysis or employing more advanced clustering algorithms can enhance accuracy and yield more profound insights. The use of interactive visualization techniques is also recommended to improve the interpretation of framing analysis results.

Acknowledgments. The authors gratefully acknowledge the Directorate of Research, Technology, and Community Service, Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology, for funding support through the 2024 Fiscal Year Beginner Lecturer Research program. Appreciation is also extended to the Directorate of Research and Community Service, Universitas Budi Luhur, for institutional support.

References

- Anggoro, A. D., Puspitasari, F. A., & Som, A. P. M. (2023). Robert Entman's Framing Analysis: Female Representation in 2024 Presidential Candidates on *Republika.com* and *Sindonews.com*. *Komunikator*, 15(2), 211–224. <https://doi.org/10.18196/jkm.19247>
- Anita, A., & Aliando, J. (2022). Pengelompokan Berita Kesehatan Pada Sosial Media Twitter Dengan Metode K-Means Clustering. *Ensiklopedia of Journal*, 4(3), 116–124. <https://doi.org/https://doi.org/10.33559/eoj.v4i3.877>
- Ashfaq, M., Shahid, N.-A., & Zubair, J. (2022). New Media and Political Protest: Framing Analysis of News Content on Twitter. *Asian Journal of Comparative Politics*, 7(4), 1177–1189. <https://doi.org/10.1177/20578911211067317>

- Boer, K. M., Pratiwi, M. R., & Muna, N. (2020). Analisis Framing Pemberitaan Generasi Mile-nial dan Pemerintah Terkait Covid-19 di Media Online. *Communicatus: Jurnal Ilmu Komunikasi*, 4(1), 85–104. <https://doi.org/10.15575/cjik.v4i1.8277>
- Chicco, D., Campagner, A., Spagnolo, A., Ciucci, D., & Jurman, G. (2025). The Silhouette co-efficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters. *PeerJ Computer Science*, 11, e3309. <https://doi.org/10.7717/peerj-cs.3309>
- Disayiram, N., & Rupasingha, R. A. H. M. (2022). A Comparative Study of Clustering English News Articles Using Clustering Algorithms. *2022 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 108–113. <https://doi.org/10.1109/SCSE56529.2022.9905210>
- Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Firman Sodik, S., Desena, W., & Wibowo, A. (2022). Penerapan Algoritma Stemming Nazief & Adriani Pada Proses Klasterisasi Berita Berdasarkan Tematik Pada Laman (Web) Direktorat Jenderal HAM Menggunakan Rapidminer. *Syntax : Jurnal Informatika*, 11(02), 10–21. <https://doi.org/10.35706/syji.v11i02.7192>
- Hamami, F., & Dahlan, I. A. (2024). Analyzing News Clustering with K-Means Algorithm: A Segmentation Perspective. *2024 4th International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 42–46. <https://doi.org/10.1109/ICE3IS62977.2024.10775603>
- Hasan, S. Al, Ruiqin, W., & Hussain, M. G. (2022). Clustering Analysis of Bangla News Articles with TF-IDF & CV Using Mini-Batch K-Means and K-Means. *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, 17–22. <https://doi.org/10.1109/CyberneticsCom55287.2022.9865339>
- Hassan, B. A., Tayfor, N. B., Hassan, A. A., Ahmed, A. M., Rashid, T. A., & Abdalla, N. N. (2024). From A-to-Z Review of Clustering Validation Indices. *Neurocomputing*, 601, 128198. <https://doi.org/10.1016/j.neucom.2024.128198>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances In the Era of Big Data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>
- Ikotun, A. M., Habyarimana, F., & Ezugwu, A. E. (2025). Benchmarking Validity Indices For Evolutionary K-Means Clustering Performance. *Scientific Reports*, 15(1), 21842. <https://doi.org/10.1038/s41598-025-08473-6>
- Jo, T. (2019). *Text Mining Concept, Implementation, and Big Data Challenge* (Vol. 45). Springer International Publishing. <https://doi.org/10.1007/978-3-319-91815-0>
- Khan, I. K., Daud, H. B., Zainuddin, N. B., Sockalingam, R., Abdussamad, A., Museeb, A., & Inayat, A. (2024). Addressing Limitations Of The K-Means Clustering Algorithm: Outliers, Non-Spherical Data, And Optimal Cluster Selection. *AIMS Mathematics*, 9(9), 25070–25097. <https://doi.org/10.3934/math.20241222>
- Khder, M. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and Its Applications*, 13(3), 145–168. <https://doi.org/10.15849/IJASCA.211128.11>

- Li, H., Sugasawa, S., & Katayama, S. (2024). Adaptively Robust and Sparse K-means Clustering. *Transactions on Machine Learning Research*. <https://doi.org/10.48550/arXiv.2407.06945>
- Semoglou, A., Likas, A., & Pavlopoulos, J. (2025). *Silhouette-Guided Instance-Weighted k-means*. <https://doi.org/10.48550/arXiv.2506.12878>
- Simanjuntak, H. T. A., Silaban, P. E. P., Manurung, J. K. S., & Sormin, V. H. (2023). Klasterisasi Berita Bahasa Indonesia Dengan Menggunakan K-Means Dan Word Embedding. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(3), 641–652. <https://doi.org/10.25126/jtiik.20231026468>
- Widaningrum, I., Mustikasari, D., Arifin, R., Tsaqila, S. L., & Fatmawati, D. (2022). Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen. *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK)*, 145–149.
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–40.
- Yazmi, R., Mala, A., Hasnan Afisah, F., Radika Sari, F., Hikam, I., Rendy Maulana, M., Rahima Orvala, A., & Ilmu Politik, P. (2024). Diskursus Politik Media Massa Indonesia: Analisis Framing Berita Pemilu 2024 dalam Surat Kabar Nasional Kompas. *Jurnal Majemuk*, 3(4), 704–715.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

