# Comparative Topic Modeling of Indian Elections: A Decadal Analysis of Media Coverage

Deval Patel
Radboud Universiteit
Nijmegen, Netherland
deval.Patel@ru.nl

## ABSTRACT

Elections significantly influence a country's political landscape, with media playing a pivotal role in shaping public discourse. This study analyzes media coverage of the 2014 and 2024 Indian general elections using Non-Negative Matrix Factorization (NMF) to uncover thematic differences across a decade. Persistent themes like economic development were identified, alongside new issues in 2024, including unemployment, caste census, and farmers' concerns. Validation against election manifestos and Wikipedia articles confirmed the relevance of generated topics, though certain manifesto priorities were underrepresented. The findings highlight how media narratives adapt to changing political contexts, offering insights into the interplay between media and democracy.

## 1 INTRODUCTION

Media coverage during elections shapes public perceptions and influences political discourse. This study examines media narratives surrounding the 2014 and 2024 Indian general elections to understand thematic shifts over a decade. Articles from *The Economic Times* were analyzed using Non-Negative Matrix Factorization (NMF), a method well-suited for identifying interpretable and non-overlapping topics from text data.

The findings highlight both persistent and evolving themes, such as the centrality of economic development and the emergence of issues like unemployment, caste census, and farmers' welfare in 2024. By validating the results against external sources, including election manifestos and Wikipedia articles, the study underscores the media's evolving role in reflecting societal and political priorities. This research contributes to understanding how media narratives adapt to changing electoral contexts.

## 2 RELATED WORK

Topic modeling has emerged as a foundational tool for analyzing large text corpora, offering insights into the underlying themes of textual data. The technique has been applied across diverse domains, including social media analysis, political discourse, and media studies, owing to its versatility in uncovering latent patterns.

### 2.1 Latent Dirichlet Allocation and Its Limitations

Latent Dirichlet Allocation (LDA) is among the most widely used topic modeling techniques. LDA employs a generative probabilistic model to discover topics as distributions over words and documents as mixtures of these topics. Jelodar et al. (2019) conducted a comprehensive survey on LDA, highlighting its applications across domains such as text summarization, sentiment analysis, and recommendation systems [8]. However, LDA's reliance on probabilistic assumptions and hyperparameter tuning often complicates interpretability, particularly for small or niche datasets. O'Callaghan et al. (2015) critiqued LDA for occasionally generating topics with high redundancy and low coherence, especially when the dataset includes overlapping or context-specific terms [10].

To address these challenges, matrix decomposition techniques like Non-Negative Matrix Factorization (NMF) have gained prominence for their interpretability and simplicity.

### 2.2 Non-Negative Matrix Factorization for Topic Modeling

NMF decomposes the document-term matrix into two non-negative matrices, representing document-topic and topic-term associations. Unlike LDA, NMF's non-negativity constraints ensure additive representations, making the topics more interpretable. Pattnaik (2020) demonstrated NMF's effectiveness in generating coherent topics from smaller datasets, particularly when preprocessing steps such as stopword removal and lemmatization are rigorously applied [13]. Similarly, O'Callaghan et al. (2015) found that NMF outperforms LDA in generating generalizable topics for niche corpora [10].

Egger and Yu (2022) extended the comparative analysis of topic modeling techniques by evaluating LDA, NMF, and modern approaches like BERTopic and Top2Vec [7]. Their findings underscored NMF's robustness in structured datasets, where it consistently produced high coherence scores. Furthermore, Maier et al. (2020) explored the effects of document sampling and vocabulary pruning on topic modeling results, demonstrating that preprocessing decisions significantly influence the quality of NMF-generated topics [9].

### 2.3 Advanced Topic Modeling Techniques

Recent advancements in topic modeling include neural and hybrid approaches, such as BERTopic and Top2Vec. These methods leverage contextual embeddings from transformer models, enabling them to capture semantic nuances that traditional techniques like LDA and NMF often overlook. Egger and Yu (2022) highlighted BERTopic's ability to dynamically adjust the number of topics, offering greater flexibility in real-time applications [7].


*TxMM'24-'25, Radboud University, Nijmegen, Netherlands*

### 2.4 The Role of Preprocessing in Topic Modeling

Effective preprocessing is a critical determinant of topic modeling success. Maier et al. (2020) demonstrated that vocabulary pruning and document sampling significantly affect coherence scores and topic quality [9]. Expanding contractions, removing high-frequency non-informative terms, and lemmatizing words are common practices that enhance the interpretability of generated topics [9, 13]. This study adopts a similar approach, iteratively refining the dataset to exclude names of politicians and political parties, which initially dominated the generated topics and hindered the identification of substantive issues.

### 2.5 Gaps in Literature Addressed by This Study

While existing research has extensively evaluated topic modeling techniques, few studies have explored the evolution of media narratives across multiple election cycles especially Indian general elections. This study addresses this gap by analyzing media articles from two distinct election periods (2014 and 2024), highlighting both persistent and emerging themes. By validating the generated topics against authoritative sources, such as manifestos and Wikipedia articles, this research contributes to understanding the interplay between media narratives, political priorities, and societal concerns.

## 3 APPROACH

The methodology involves three main steps: data collection, topic modeling using NMF, and validation against external sources.

### 3.1 Data Collection

Data was sourced from the archives of *The Economic Times*, focusing on articles published during the months leading up to the 2014 and 2024 Indian general elections. The 2014 dataset included 407 articles from February and March, while the 2024 dataset comprised 277 articles from February through early April.

Initial attempts to automate data collection using tools like Beautiful Soup and Newspaper3k were unsuccessful due to incomplete scraping and challenges in filtering relevant articles. Articles scraped using these libraries were often incomplete, ending abruptly with lines such as:

`"Read more at:` https://economictimes.indiatimes.com/news/politics-and-nation"

As a result, the entire data collection process had to be performed manually. Articles were carefully selected at intervals of four days to ensure comprehensive coverage while maintaining a manageable dataset size. This approach ensured that significant events and developments were captured without redundancy, aligning with the study's objectives.

### 3.2 Topic Modeling

Topic modeling was conducted using Non-Negative Matrix Factorization (NMF), a matrix decomposition technique widely recognized for its interpretability and efficacy in extracting coherent topics[1, 10, 13]. Unlike probabilistic approaches like Latent Dirichlet Allocation (LDA), NMF decomposes the document-term matrix into two non-negative matrices, representing the association of documents with topics and topics with terms, respectively [2, 8, 10]. This additivity ensures that the topics are interpretable and align with the intuitive understanding of themes present in the dataset. The objective of applying NMF in this study was to discover themes in media articles and analyze how these topics evolved over two election cycles.

*3.2.1 Exploratory Data Analysis and Preprocessing.* Exploratory Data Analysis (EDA) was performed to understand the characteristics of the dataset and guide preprocessing decisions. Key observations included:

- **Word Frequencies:** Most of the highly frequent terms were stopwords and proper nouns. Highly frequent proper nouns included names of prominent politicians and political parties. While these terms were initially retained, their dominance in the generated topics necessitated retrospective adjustments (explained below)[9].
- **Special Characters and contractions:** All the articles contained special characters and contractions that did not contribute meaningfully to the analysis. These were removed during preprocessing.

Based on these findings, the preprocessing pipeline was designed to include the following :

- Expanding contractions.
- Removal of special characters, numbers, and extra spaces.
- Removal of single characters that occur naturally or as a result of expanding contractions.
- Converting text to lower case
- Tokenization
- Removal of stopwords and special words
- Lemmatization

This ensured consistency and reduced noise. Additionally, frequent but non-informative terms identified during EDA, such as politician and party names, were retrospectively excluded from the corpus to improve the relevance of the generated topics[9].

*3.2.2 Challenges and Adjustments.* The initial application of NMF generated topics that were dominated by the names of politicians and political parties. While these terms are closely associated with elections, their prevalence masked the underlying issues discussed in the articles. For instance, instead of identifying thematic topics such as economic development, or caste census, the topics merely reflected the mentions of political figures, providing limited insight into the narrative.

To address this, a list of special words, including the names of key politicians and parties, was excluded during preprocessing[9]. This adjustment allowed the NMF model to focus on more substantive terms, resulting in topics that better represented the issues being discussed, such as development, corruption, and unemployment.

*3.2.3 Implementation and Results.* The preprocessed dataset was vectorized using TF-IDF Vectorizer (Term Frequency-Inverse Document Frequency) to create a sparse matrix, which served as input for the NMF model. The number of topics was determined empirically by evaluating reconstruction error and topic coherence scores across a range of values. The final model produced interpretable and coherent topics, each represented by its top 10 contributing terms.

The refined topics captured key election-related issues, providing insights into the priorities and narratives emphasized in media coverage. The improved results were also a result of removing some special words, which in itself was a big challenge as removing a large number of proper nouns resulted in worse results. So, a balance had to be found so that generated results would be relevant.

By addressing the initial challenges and refining the preprocessing steps, the topic modeling approach effectively uncovered the themes in the dataset, aligning with the objectives of this study.

## 3.3 Validation

The quality of the topics generated by the NMF model was validated using both quantitative and qualitative measures. These measures ensured that the identified topics were both mathematically robust and semantically meaningful.

*Reconstruction Error.* Reconstruction error was used as a key metric to assess the model's fit. This error measures the difference between the original TF-IDF matrix ($V$) and its approximation ($W \times H$), where $W$ is the document-topic matrix and $H$ is the topic-term matrix. A lower reconstruction error indicates a better approximation and suggests that the selected number of topics adequately captures the underlying structure of the data.

*Coherence Analysis.* To ensure the semantic interpretability of the generated topics, coherence analysis was performed using the `c_v` coherence metric [10]. This metric evaluates how closely related the top terms within a topic are, based on their co-occurrence patterns in the dataset. For each candidate number of topics, the NMF model was fitted, and the coherence scores were calculated. These coherence scores, calculated across a range of topic numbers, were used to identify the optimal number of topics for the study.

*Qualitative Validation.* To validate the relevance of the generated topics, a qualitative approach was employed. The top terms associated with each topic were compared against publicly available resources, including Wikipedia articles and the election manifestos of the two largest political parties in India, the Bharatiya Janata Party (BJP) and the Indian National Congress. By aligning the generated topics with external references each generated topic was given a meaningful topic name. The qualitative validation ensured that the topics reflected meaningful and contextually accurate narratives. This process provided an additional layer of confidence in the interpretability and reliability of the generated topics.

# 4 RESULTS

## 4.1 Introduction

This section presents the key findings from the topic modeling analysis of media articles covering the 2014 and 2024 Indian general elections. The results include the identification of dominant topics and their evolution across the two election cycles, quantitative evaluation of the topic modeling approach, and validation against external sources. A detailed interpretation of these findings is provided in the Discussion section.

## 4.2 Key Topics and Their Evolution

The application of NMF revealed distinct topics for the 2014 and 2024 elections, as shown in Tables 1 and 2. Persistent themes such as economic development were observed across both cycles, while newer issues like unemployment, caste census, and farmers' welfare gained prominence in 2024. Controversies played a major role in both cycles, with the 2014 election dominated by discussions on the Telangana State separation and the Gujarat riots, whereas the 2024 cycle focused on electoral bonds and the Citizenship Amendment Act (CAA). These findings underscore the evolution of societal and political priorities over the decade, with media narratives adapting to the shifting contexts of each election.

## 4.3 Quantitative Evaluation

Reconstruction error and coherence scores validated the quality of the NMF model. Coherence analysis determined the optimal number of topics, ensuring that the generated themes were interpretable and consistent. These metrics, alongside manual validation, confirmed the reliability of the findings. Table 3 presents the reconstruction error and coherence score for the optimal number of topics selected. Detailed visualizations of the reconstruction error and coherence scores are provided in the Appendix (Figures 1,2, 3, and 4). Further analysis and implications of these scores are discussed in the Discussion section.

## 4.4 Validation Outcomes

The topics were validated against external sources, including Wikipedia articles of both elections and the election manifestos of the Bharatiya Janata Party (BJP) and the Indian National Congress (INC), two of the biggest political parties of India.

The qualitative validation process revealed that the generated topics reflected several key themes emphasized in these authoritative sources. For instance, topics related to economic development, inflation, and corruption in 2014 were consistent with the policy priorities outlined in the BJP and INC manifestos, as well as the documented issues from the 2014 Indian general election Wikipedia page [3, 5, 11]. Similarly, mentions of inflation, farmers' issues, economic development, and the caste census in the 2024 topics aligned with the themes highlighted in the manifestos and the 2024 Indian general election Wikipedia page [4, 6, 12].

However, despite these similarities certain manifesto priorities—like women's issues and healthcare—did not surface in the topics. This suggests potential biases in media coverage or data limitations which have been further discussed in the discussion section.

# 5 DISCUSSION

This study explored the thematic focus of media coverage during the 2014 and 2024 Indian general elections using Non-Negative Matrix Factorization (NMF). The results revealed distinct themes and controversies for each election cycle, highlighting both similarities and differences in the issues emphasized. This section delves into the key findings, methodological insights, and limitations, providing a comprehensive understanding of the study's contributions.

### 5.1 Key Findings

The topics generated for the 2014 election cycle captured many of the main issues documented in manifestos and external sources, including economic development, inflation, and corruption [3, 5, 11]. However, certain themes, such as education reforms, healthcare, housing, and the youth agenda, were absent. This could be attributed to the dataset's limited size, where these issues were not sufficiently covered. Interestingly, many topics reflected the media's focus on controversies, such as the Telangana State separation, the Ashok chavan issue, Arvind Kejriwal comments on sending journalists to jail, and the Gujarat riots. This underscores the role of the media in emphasizing political accusations and disputes, often at the expense of broader policy discussions.

Similarly, the 2024 topics identified issues like unemployment, farmers' concerns, and the caste census, consistent with key manifesto points [4, 6, 12]. However, some significant themes, such as the Ram Mandir consecration, the Hindutva agenda, and women's issues, did not surface in the results. As in 2014, media coverage appeared to focus heavily on controversies, including electoral bonds, the Citizenship Amendment Act (CAA) bill, and the white paper presentation. The coherence score of 0.69 and reconstruction error of 14.68 for 2024 were comparable to the 2014 dataset, with a coherence score of 0.67. These scores validate the overall interpretability of the topics generated but highlight trade-offs in topic granularity when optimizing for coherence.

The comparative analysis revealed both persistent and evolving themes. Economic development and inflation remained central to public discourse in both elections, reflecting ongoing developmental priorities. However, the emergence of topics like the caste census and defending the constitution in 2024 indicated a shift towards addressing social issues alongside economic ones. The consistent focus on seat allocation across both cycles illustrates the importance of electoral logistics in Indian politics.

### 5.2 Methodological Insights

The use of NMF proved effective for generating interpretable topics from media articles. Iterative preprocessing, particularly the exclusion of a "special words" set comprising names of politicians and political parties, significantly improved the quality of results by shifting the focus from individuals to broader issues. However, this process was time-consuming, as it involved trial-and-error adjustments over multiple iterations. Different word sets yielded varying results, changing both the topics and the optimal number of topics, suggesting that the set used may not have been the most optimal.

Coherence analysis using the `c_v` metric was instrumental in determining the appropriate number of topics. Balancing coherence scores with reconstruction error ensured that the selected topics were both interpretable and representative of the dataset. The slightly higher reconstruction error observed in this study can be attributed to the inherent variability and complexity of election-related media narratives, which often involve overlapping themes and context-specific terms that are challenging for matrix factorization techniques to capture fully. External validation using Wikipedia articles and manifestos further reinforced the reliability of the results.

### 5.3 Limitations

This study is subject to several limitations that may have influenced the findings. The dataset was relatively small, comprising articles from a single news outlet over a span of approximately two months before each election. This limited scope may have excluded key issues, particularly those underrepresented in the media or discussed on non-selected days. The manual data collection process, although careful, might have inadvertently introduced biases, as the inclusion or exclusion of certain articles could affect the overall results.

The preprocessing phase relied heavily on the creation of a "special words" set through iterative adjustments. While this improved the interpretability of topics, it introduced variability and required substantial effort. Additionally, the choice of NMF, while effective for this task, might not fully capture nuanced themes. Use of advanced models like BERTopic or Top2Vec might give more nuanced themes for the same data[7].

### 5.4 Broader Implications

The findings offer valuable insights into the evolution of media narratives during elections, highlighting the interplay between societal priorities, political agendas, and media focus. The persistent emphasis on controversies suggests that media coverage may prioritize sensationalism over substantive issues, potentially shaping public perceptions of elections. The absence of certain manifesto themes in the results further raises questions about the comprehensiveness of media reporting.

Future research could address these limitations by incorporating larger, more diverse datasets, including social media content and articles from regional outlets. Exploring advanced topic modeling techniques could also provide richer, more nuanced insights into electoral narratives.

## 6 CONCLUSION AND OUTLOOK

This study employed Non-Negative Matrix Factorization (NMF) to analyze media narratives surrounding the 2014 and 2024 Indian general elections, uncovering key themes and shifts in public and political discourse over the decade. By examining articles from *The Economic Times*, the research identified persistent issues such as economic development and inflation, alongside emerging topics in 2024, including unemployment, the caste census, and farmers' concerns. These findings highlight the dynamic nature of societal priorities and media coverage during election cycles.

The results also underscored the role of controversies in shaping election narratives, with the focus evolving from the 2002 Gujarat riots and Telangana State separation in 2014 to electoral bonds and the Citizenship Amendment Act (CAA) in 2024. While the generated topics aligned with several key issues mentioned in election manifestos and Wikipedia articles, some important themes, such as education reforms, healthcare, and women's issues, were notably absent. This points to potential biases in media reporting and the limitations of the dataset used.

Despite its contributions, this study has certain limitations. The dataset, derived from a single media source, may not fully capture the diversity of narratives presented in regional outlets or social media platforms. The iterative preprocessing approach, while effective, required substantial manual effort and could introduce variability in results. Additionally, NMF's reliance on linear relationships between terms may have restricted its ability to capture more nuanced themes.

Future research could address these challenges by incorporating larger, more diverse datasets, including regional and multilingual media sources. Advanced topic modeling techniques, such as BERTopic or neural-based models, could further enhance the granularity and interpretability of results[7, 14]. Expanding the scope to include analyses across multiple election cycles would provide a deeper understanding of how media narratives evolve over time.

This study contributes to the field of political analysis by demonstrating how topic modeling can uncover latent themes in media coverage. It highlights the importance of critically examining media narratives to understand their impact on public discourse and democratic processes. By building on these findings, future work can continue to shed light on the interplay between politics, media, and society.

## REFERENCES

[1] Stella Cherotich. 2020. Discovering Topic Modeling with NMF. https://stellacherotich.medium.com/discovering-topic-modeling-with-nmf-fe09c67d5f22. (2020).
[2] Coding Club. 2020. Topic Modeling in Python: An Introduction. https://ourcodingclub.github.io/tutorials/topic-modelling-python/.
[3] Indian National Congress. 2014. Congress Election Manifesto 2014. https://www.indiaspend.com/h-library/congress-manifesto-2014.pdf
[4] Indian National Congress. 2024. Congress Election Manifesto 2024. https://manifesto.inc.in/en/introduction/
[5] Wikipedia contributors. 2014. 2014 Lok Sabha Elections. https://en.wikipedia.org/wiki/2014_Indian_general_election#Issues.
[6] Wikipedia contributors. 2024. 2024 Lok Sabha Elections. https://en.wikipedia.org/wiki/2024_Indian_general_election#Major_election_issues.
[7] Roman Egger and Joanne Yu. 2022. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology* 7 (2022). https://doi.org/10.3389/fsoc.2022.886498
[8] Hamed Jelodar, Yong Wang, Chengqi Yuan, et al. 2019. Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. *Multimed Tools Appl* 78 (2019), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4
[9] Daniel Maier, Andreas Niekler, Gregor Wiedemann, and David Stoltenberg. 2020. How Document Sampling and Vocabulary Pruning Affect the Results of Topic Models. *Computational Communication Research* 2, 2 (2020), 139–152. https://computationalcommunication.org/ccr/article/view/32
[10] Derek O'Callaghan, Derek Greene, Joe Carthy, and Pádraig Cunningham. 2015. An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications* 42, 13 (2015), 5645–5657. https://doi.org/10.1016/j.eswa.2015.02.055
[11] Bharatiya Janata Party. 2014. BJP Election Manifesto 2014. https://www.bjp.org/bjp-manifesto-2014
[12] Bharatiya Janata Party. 2024. BJP Election Manifesto 2024. https://www.bjp.org/bjp-manifesto-2024
[13] Priyanka P. Pattnaik. 2020. Topic Modeling with Non-Negative Matrix Factorization (NMF). https://medium.com/analytics-vidhya/topic-modeling-with-non-negative-matrix-factorization-nmf-3caf3a6bb6da.
[14] Jianyu Wang and Xiao-Lei Zhang. 2023. Deep NMF topic modeling. *Neurocomputing* 515 (2023), 157–173. https://doi.org/10.1016/j.neucom.2022.10.002

## A WORK REPORT

The project began with an attempt to scrape articles from news websites using pre-existing libraries. However, this approach proved unsuccessful due to technical limitations. Subsequently, articles were manually collected from the archives of *The Economic Times*. Initially, 220 articles from February 2014 were gathered, and the Non-Negative Matrix Factorization (NMF) pipeline was developed. A trial run revealed that the results were heavily dominated by the names of political parties and politicians. To address this issue, a "special words set" was introduced into the preprocessing pipeline. Although this improved the results, they were still suboptimal, prompting iterative refinements to the special words set.

Results were still suboptimal so, additional articles from March 2014 were collected, resulting in a total of 407 articles for the 2014 election cycle. The NMF pipeline was then re-run on this expanded dataset, with further iterative improvements to the special words set. The final topics were manually validated to ensure their relevance and coherence.

For the 2024 election cycle, articles from February and March 2024 were collected, yielding a dataset of 277 articles. The NMF pipeline was applied to this dataset, and the special words set was once again iteratively refined. The generated topics were manually validated, ensuring consistency with the research objectives.

Throughout the research process, notes were kept to capture every detail and address all aspects of the study in the paper. The writing phase involved multiple iterations to refine and improve the paper. ChatGPT was used for coding assistance, language enhancement, and improving overall coherence and readability of the research paper. Care was taken to ensure that no factual content was introduced by ChatGPT and that the study's findings and analysis remained entirely original. Drafts were submitted to ChatGPT for grammar correction and polishing, ensuring that the final version was polished and professional.

## B FIGURES

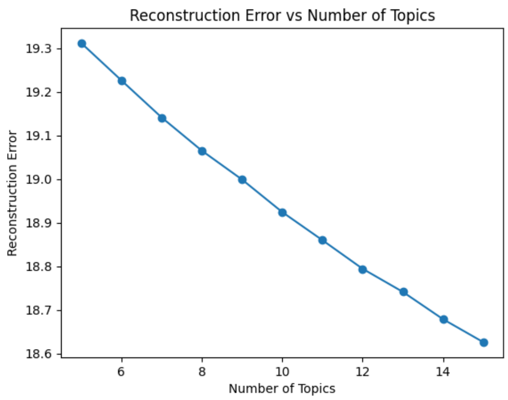


**Figure 1: Reconstruction Error vs. Number of Topics for the 2014 Dataset.**

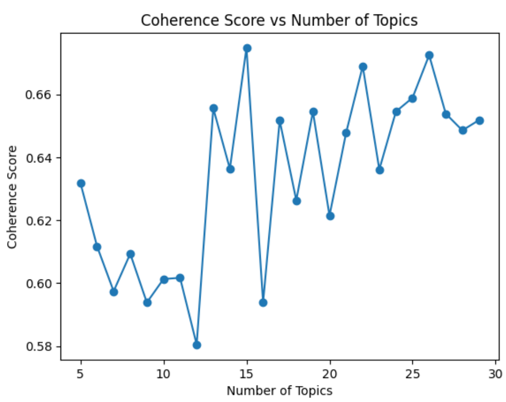


**Figure 2: Coherence Score vs. Number of Topics for the 2014 Dataset.**

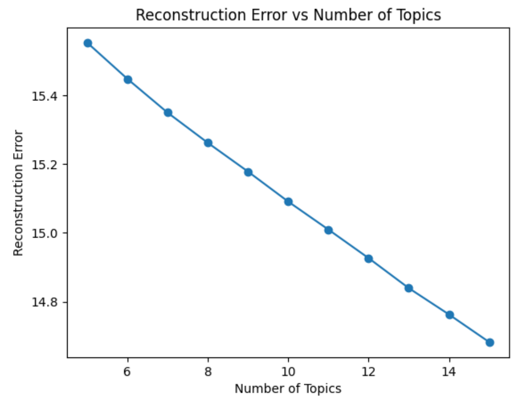


**Figure 3: Reconstruction Error vs. Number of Topics for the 2024 Dataset.**

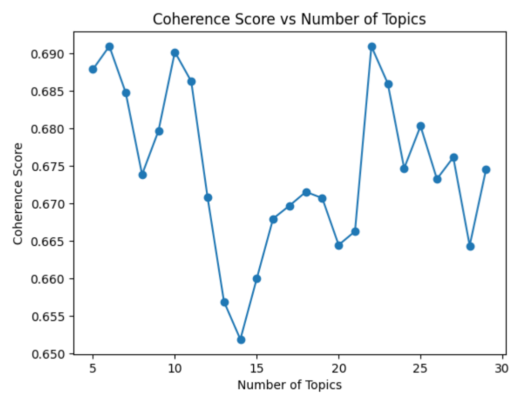


**Figure 4: Coherence Score vs. Number of Topics for the 2024 Dataset.**

## C TABLES

**Table 1: Topics Identified in the 2014 Election Articles and Their Top Terms.**

| Topic | Top Terms |
|---|---|
| Seat Allocation | seat, candidate, sabha, lok, election, poll, varanasi, leader, campaign, ticket |
| Economy | budget, interim, government, economy, finance, raja, fiscal, certain, going, election |
| Telangana Issue | telangana, seemandhra, andhra, region, state, tdp, reddy, pradesh, mp, jagan |
| Jan Lokpal Bill | lokpal, swaraj, jurist, panel, rao, appointment, opposition, selection, sushma, member |
| Gujarat development questions | gujarat, state, people, development, minister, riot, country, rally, chief, prime |
| Inflation and economy growth | manifesto, growth, cent, sector, india, policy, economic, tax, inflation, job |
| Arvind Kejriwal comments on sending journalists to jail | medium, remark, electronic, jail, leader, social, crush, sold, comment, journalist |
| Party Politics and Alliances | alliance, tamil, nadu, party, ally, seat, poll, chennai, pmk, touch |
| Seemandhara Issue | bihar, special, status, category, bandh, state, betrayal, jdyou, seemandhra, demand |
| Accusations | rss, patel, sardar, mahatma, assassination, gandhi, gandhiji, secretary, ec, killed |
| Corruption | corrupt, corruption, list, assam, politics, politician, minister, gogoi, allegation, chief |
| Black money | hard, money, black, fm, working, upas, paralysis, pension, policy, chidambarams |
| Politicians moving between parties | party, patnaik, left, formation, jdyou, noncongress, nonbjp, alternative, odisha, bjd |
| Jan Lokpal Bill | binny, government, jan, support, lokpal, delhi, assembly, power, mlas, mla |
| Ashok chavan Issue | cbi, director, statement, mastered, art, independent, sinha, manipulation, chavan, pliability |

**Table 2: Topics Identified in the 2024 Election Articles and Their Top Terms.**

| Topic | Top Terms |
|---|---|
| Seat Allocation | congress karnataka seat bjp party chhindwara sabha candidate mp lok |
| Electoral Bonds | bond electoral crore donor donation company court profit donated sbi |
| CAA Issue | caa citizenship bengal matua rule amendment bangladesh tmc act law |
| Defending the Constitution | constitution rally democracy minister india country opposition prime people |
| Katchatheevu Issue | dmk tamil nadu katchatheevu lanka aiadmk sri stalin island minister |
| Topic 6 | delhi aap punjab candidate swaraj lok sabha mann seat aam |
| Caste census | manifesto nyay congress guarantee yatra caste party promise census promised |
| White Paper contoversy and economy | paper upa white sitharaman government economy economic crisis finance budget |
| Seat Allocation | party bjp state strength poll weakness seat opportunity threat assembly |
| Inflation, Farmers Issues and youth unemployment | farmer government msp price post support border youth life bharat |

**Table 3: Reconstruction Error and Coherence Score**

| Data | Topics Selected | Reconstruction Error | Coherence Score |
|---|---|---|---|
| 2014 data | 15 | 18.626045 | 0.674772 |
| 2024 data | 10 | 14.682873 | 0.690196 |