Sentiment Forecasting of Political Party: Analyzing Public
and News Media Sentiment Over Time

Leveraging self-supervised sentiment models and temporal forecasting to track political sentiment in public discourse and media influence

Under the guidance of
Prof. Saurish Dasgupta
Course: Neural Network & Deep Learning
M.Sc. Data Science
Dhirubhai Ambani University

Team members:
Milan Nagvadiya (202418035)
Krisha Sompura (202418026)
Sheetal Jain (202418051)
Yashraj Sinh (202418064)

Abstract

This project forecasts sentiment trends toward the Bharatiya Janata Party (BJP) from 2013 to 2018, analyzing public tweets and news articles to understand political sentiment dynamics during events like the 2014 Indian General Elections. Using self-supervised sentiment analysis, time-series forecasting, and comparative media analysis, we uncover patterns in public and institutional narratives, offering insights into their influence on political discourse.

1 Introduction

Political sentiment shapes public discourse in the digital age, influencing how parties like the Bharatiya Janata Party (BJP) engage with citizens, especially during transformative events such as the 2014 Indian General Elections, 2016 Demonetization, and 2017 GST implementation. Social media platforms like Twitter provide a real-time lens into public opinion, while news outlets offer curated perspectives that may reflect editorial biases. This project examines sentiment trends toward the BJP from 2013 to 2018, leveraging both tweets and news articles to explore the interplay between public and media narratives. By analyzing these dual perspectives, we aim to understand how sentiment evolves and impacts political strategies, providing a foundation for more informed decision-making in a rapidly evolving political landscape

2 Problem Definition

The lack of labeled sentiment data poses a significant challenge for analyzing political sentiment toward the Bharatiya Janata Party (BJP) from 2013 to 2018, a period marked by key events. Traditional supervised learning methods are impractical, necessitating a self-supervised approach to generate continuous sentiment scores from unlabeled tweets. Additionally, the noisy, unstructured nature of tweets and the contextual nuances in news articles require robust preprocessing and modeling techniques to capture sentiment accurately. This project addresses these challenges with two objectives: first, to develop a sentiment analysis pipeline using pretrained models (CardiffNLP, BERTweet, VADER) for tweets and a Bidirectional LSTM with BERT embeddings for news articles; second, to forecast sentiment trends using time-series methods, proposing the pRT+ model with Multi-Head LSTMs and attention mechanisms for tweets, while ultimately adopting the Informer model for its superior performance. By integrating engagement metrics (likes, retweets) and comparing public and media sentiment, we aim to identify patterns, fluctuations, and biases, delivering an innovative framework to guide strategic communication in political discourse.

The objective is to forecast sentiment trends toward the BJP using tweets, while also predicting sentiment scores for news articles from a single outlet during the same period. We then analyze the interplay between public sentiment (from tweets) and media sentiment (from news articles).

Tweet Representation

Each tweet is represented as $\text{Tweet}_j = (d_j, r_j, e_j)$, where:

- d_j: posting date of the j-th tweet.
- r_j: raw text content.
- $e_j = (R_j, L_j)$: engagement metrics (retweets R_j, likes L_j).
- $j = 1, 2, \ldots, T$, where T is the total number of tweets.

News Article Representation

Each news article is represented as $\text{Article}_k = (d_k, a_k)$, where:

- d_k: publication date of the k-th article.
- a_k: article text content.
- $k = 1, 2, \ldots, N$, where N is the total number of articles.

Sentiment Scoring

- For tweets, compute a sentiment score $S_j \in [0, 1]$ for Tweet_j:
  $S_j = f_{\text{score}}(r_j)$, where f_{score} uses pretrained models (CardiffNLP, BERTweet, VADER).
- For news articles, predict a sentiment score $S_k \in [0, 1]$ for Article_k:
  $S_k = g_{\text{score}}(a_k)$, using a Bidirectional LSTM with BERT embeddings.

Forecasting Objective (Tweets)

Forecast future tweet sentiment trends:

$$\hat{S}_{\text{tweet}}(t + 1) = h_{\text{forecast}}(S_{\text{tweet}}(t), e_t)$$

where:

- $e_t = (R_t, L_t)$: aggregated engagement metrics at time t.
- h_{forecast}: forecasting model (e.g., pRT+ with Multi-Head LSTMs and attention).

Prediction of sentiment score for News

Precdict sentiment score using bert/glove embeding:

$$\hat{S}_{\text{news}}(t) = h_{\text{predict}}(S_{\text{news headline}}(t))$$

where:

- h_{predict}: Score prediction model (e.g.glove1/bert emmbeding with bidirectional LSTM model).

Analysis Objective

Compare the forecasted tweet sentiment $\hat{S}_{\text{tweet}}(t)$ and predicted news sentiment $S_{\text{news}}(t)$ to identify patterns and biases across public and media narratives.

3 Previous Work

Tweet sentiment analysis has primarily focused on classification tasks, using tools like VADER to categorize sentiment as positive, negative, or neutral. News sentiment analysis, often employing static embeddings like GloVe, has been widely used to forecast election results and assess media bias during political events.

3.1 Loopholes in Previous Work

Despite advancements, prior work has key limitations:

- **Weak Contextual Understanding:** VADER struggles with sarcasm, slang, and emojis in tweets, while GloVe misses nuances like sarcasm in news, leading to inaccurate sentiment scoring.
- **Poor Temporal Modeling:** Models like ARIMA fail to capture non-linear sentiment shifts during viral events, such as the 2016 Demonetization.
- **Limited Engagement Use:** Engagement metrics (likes, retweets) are often underutilized, ignoring their role in shaping sentiment trends.
- **Lack of Comparative Analysis:** Few studies explore the interplay between social media and news sentiment, missing insights into their mutual influence during events.
- **Labeling Issues:** Manual labeling limits scalability and introduces bias, with self-supervised methods underused for continuous sentiment scoring.

3.2 Our Contribution

Our project addresses existing gaps in sentiment analysis by employing a multifaceted approach. First, we utilize self-supervised scoring mechanisms with pretrained language models to achieve robust and scalable sentiment labeling. To effectively capture temporal dynamics and user engagement patterns in tweets, we introduce a novel model, pRT+. For news articles, we leverage a Bidirectional LSTM architecture integrated with BERT embeddings to better capture contextual subtleties and improve sentiment interpretation. Additionally, we conduct a comprehensive comparative analysis between tweet and news sentiments to uncover their interdependencies and broader implications for public discourse analysis.

4 Methodology

Our approach is divided into two distinct pipelines: one for processing and forecasting sentiment from political public tweets, and another for prediction of sentiment in news articles towards party.

4.1 Tweet Sentiment Score Forecasting

4.1.1 Tweet Preprocessing

Raw tweet data is standardized using a custom preprocessing module. This includes normalizing informal terms, mapping slang to standardized entities via a political synonyms dictionary, cleaning hashtags, URLs, and emojis with Ekphrasis, and performing additional processing such as emoji-to-text conversion, lemmatization, stopword removal, URL removal, and case normalization.

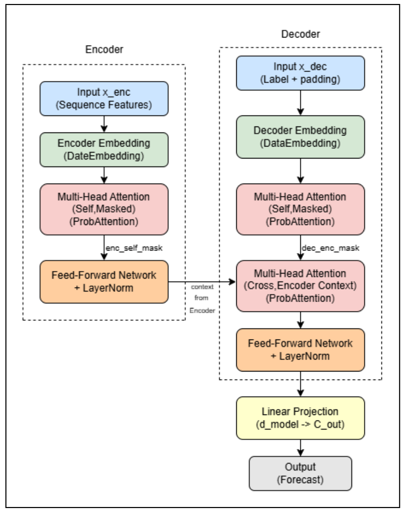

Figure 1: Proposed architecture for forecasting tweet sentiment trends.

4.1.2 Sentiment Scoring

Sentiment scores are assigned to pre-processed tweets using three pretrained models: **CardiffNLP**, **BERTweet**, and **VADER**.

4.1.3 Model Selection via Consensus Metrics

The optimal model is determined by evaluating sentiment scores across various models.

Mean Sentiment Score ($\mathbf{S_{mean}}$): For each tweet T_j, the mean sentiment score is computed as:

$$S_{\text{mean}}(T_j) = \frac{1}{3} \sum M_i(T_j)$$

where $M_i(T_j)$ denotes the sentiment score provided by model i for tweet T_j.

1. **Mean Absolute Error (MAE):** Measures the average absolute difference between the model's sentiment score and the mean sentiment score.
2. **Root Mean Squared Error (RMSE):** Quantifies the square root of the average squared difference between the model's sentiment score and the mean sentiment score.

Additional Metrics: In addition to MAE and RMSE, **Euclidean Distance**, **Pearson Correlation**, and **Cosine Similarity** are used for comprehensive evaluation.

Model Ranking: The **Mean Reciprocal Rank (MRR)** is computed based on the ranking of each model's prediction for each tweet.

Based on these consensus metrics, **CardiffNLP** achieved the highest MRR and was selected as the optimal model.

4.1.4 Forecasting Tweet Sentiment

For forecasting sentiment score of tweets towards BJP we proposed forecasting model prt+ as shown in figure 1.

Proposed architecture: Fig 1

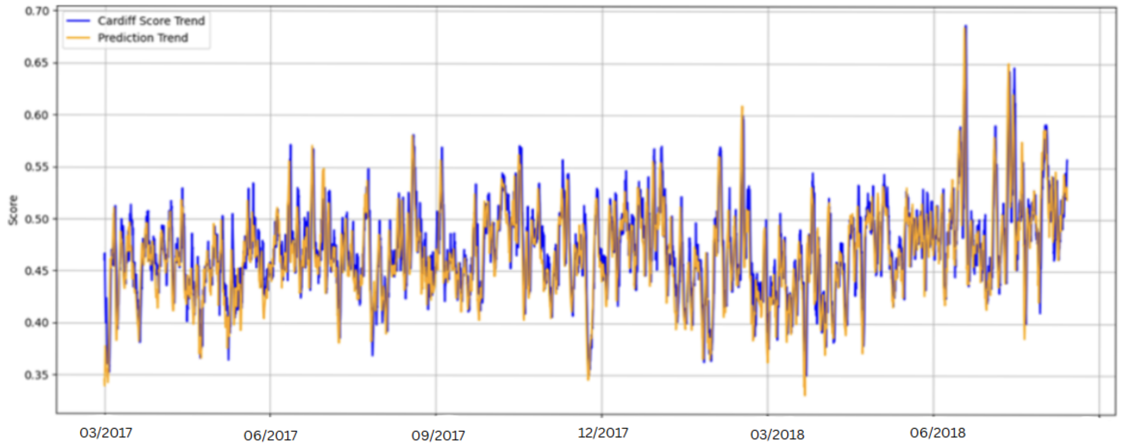

Figure 2: Cardiff Score Trend with Predicted Sentiment Trend

4.1.5 Key Architectural Features and Considerations

- **Encoder-Decoder Structure:** The model follows a standard encoder-decoder pattern, which is well-suited for sequence-to-sequence tasks.

- **Attention Mechanisms:** The use of multi-head attention is a key feature, allowing the model to selectively focus on the most relevant parts of the input sequence and the decoder's previous outputs. The use of "ProbAttention" suggests a focus on computational efficiency.

- **Masking:** Masking in the self-attention layers is essential for both the encoder and decoder to ensure proper handling of sequential data and prevent information leakage from the future.

- **Embedding Layers:** Embedding layers are used to transform the input features and labels into a dense, vector representation, which is standard practice in deep learning for sequence data.

- **Feed-Forward Networks and Layer Normalization:** These components help to improve the model's capacity and stability.

4.1.6 Performance Metrics

Metric	Value
MAE	0.045858573168516l6
MSE	0.15602150559425354
SMAPE	32.32996368408203%

Table 1: Performance metrics for the proposed model

4.1.7 Comparison with SOTA and Baseline Models

Model	SMAPE	MSE	MAE	Training Time (s)
pRTplus	**32.329964**	0.045859	0.156022	2224.1
ARIMA	35.809497	0.050881	0.169286	1530.0
GRU	32.6002459	0.044787	0.153773	256.0
RNN	200.066153	13148.665039	115.8376	1882.0
Informer_metrics	32.754615	0.045898	0.151487	1371.0
LSTM	32.764053	0.044789	0.151497	230.8

Table 2: Comparison of proposed model with baseline and state-of-the-art models

4.1.8 Discussion

The pRT+ model demonstrated strong forecasting accuracy but faced challenges with high training times and complexity. Despite these issues, pRT+ showed good generalization performance, similar to Informer, and had a stable loss function, unlike models like GRU and LSTM. Informer was ultimately preferred for its balance of accuracy, speed, and robustness in long-term political sentiment forecasting. However, pRT+ has potential, and its performance could improve with a larger training dataset, making it a strong contender for real-world applications.

4.2 Prediction of News Sentiment

4.2.1 Data Collection

To ensure a more generalized and representative dataset for sentiment analysis, we combined word-level sentiment data from global news sources with Indian news headlines, both of which were obtained from publicly available online platforms. The compiled dataset included news texts along with their corresponding sentiment score , providing a rich foundation for training and evaluation.

4.2.2 Dataset Preparation

The scraped headlines underwent preprocessing, including lowercasing, removal of punctuation, Unicode normalization, and tokenization . This process yielded a structured training dataset

$$D = \{(A_i, S_i)\}_{i=1}^{N} = \{(A_i, S_i)\}$$

where A_i represents the preprocessed headline text and $S_i \in [0, 1]$ denotes the corresponding sentiment score, assigned through a supervised deep learning model.

4.2.3 Embedding Generation

Two distinct embedding strategies were employed to represent the headline text:

- **GloVe Embeddings:** Pre-trained 100-dimensional GloVe vectors were utilized to construct an embedding matrix $E_{\text{GloVe}} \in R^{|V| \times 100}$, providing static word representations.

- **BART Embeddings:** Contextual embeddings were generated using BERT's 100-dimensional vectors, incorporating attention masks via the bert-base-uncased tokenizer to capture nuanced semantic relationships.

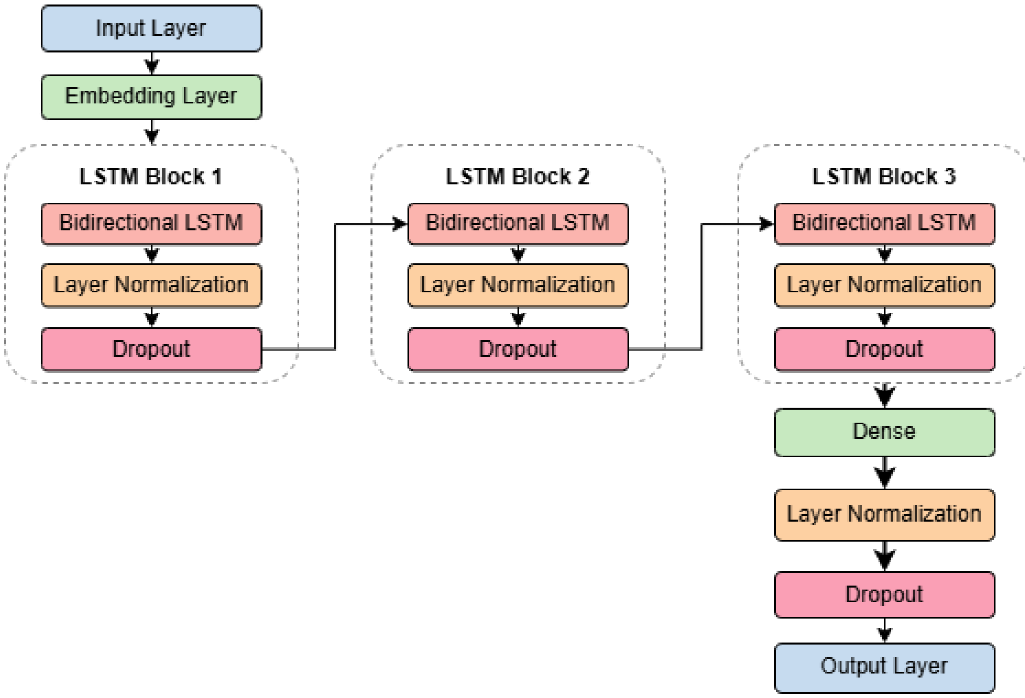

Figure 3: Bi-Directional Lstm architecture

4.2.4 Model Training

A bidirectional Long Short-Term Memory (BiLSTM) model was trained using the BERT embeddings. The architecture comprised a 100-dimensional BERT embedding layer trainable/Glove embedding layer followed by three BiLSTM layers (each with 128 units), augmented by layer normalization and a 0.2 dropout rate for regularization. The output was processed through dense layers with ReLU and sigmoid activations to produce a sentiment score $S_i \in [0, 1]$.

The dataset was partitioned into 80% training, 10% validation, and 10% testing sets. Training was conducted over 15 epochs, optimizing for Mean Squared Error (MSE) loss and tracking Mean Absolute Error (MAE) as a metric, with the Adam optimizer and early stopping to prevent overfitting.

4.2.5 Performance Evaluation

The performance of the GloVe and BERT-based models was evaluated on the validation set, yielding the following metrics:

Model	Val MSE	Val MAE
GloVe	0.035320	0.130660
BERT/BART	0.015855	0.077860

Table 3: Validation performance of GloVe and BERT-based BiLSTM models

The BERT model demonstrated superior performance, achieving a lower MSE and MAE compared to the GloVe-based approach.

4.2.6 Sentiment Scoring of Indian News Headlines

The saved BERT-based model was applied to newly scraped Indian news headlines related to the BJP. Following preprocessing consistent with the training phase, the model predicted the sentiment scores (0,1). These scores were aggregated to analyze the sentiment trajectory, providing insight into the narrative of the news channel surrounding BJP-related events, different from the sentiment dynamics of social media.

5 Time based analysis of People's and News sentiment score

This section investigates the temporal evolution of sentiment scores derived from tweets and news articles referencing the Bharatiya Janata Party (BJP). The aim is to determine whether the media influences public sentiment or responds to it—especially during key political milestones. Public reactions on platforms like Twitter are compared with editorial trends in news coverage to assess alignment or divergence. This analysis offers insight into the interplay between news narratives and people's opinions, raising questions about agenda setting, information framing, and responsiveness to mass sentiment.

5.1 2014 General Elections: Fig 4:(a)

As the BJP prepared for the 2014 elections, news channels adopted a positive tone, emphasizing Narendra Modi's leadership and anti-corruption stance. This media framing cultivated optimism, mirrored in Twitter discourse. Post-election, news sentiment peaked at 0.50, and Twitter sentiment steadily aligned, reflecting media-driven political euphoria.

5.2 2016 Demonetization: Fig 4:(b)

In November 2016, news outlets initially portrayed Demonetization as a bold move against black money, with sentiment at 0.55. Twitter sentiment initially dropped to 0.35 due to public hardships but gradually rose back to 0.50 by late 2016, suggesting media influence over public perception.

5.3 2017 GST Rollout: Fig 4:(c)

Post-GST implementation in July 2017, media coverage turned critical, citing economic slowdown and logistical issues, reflected in a 0.35 sentiment score. However, Twitter sentiment remained stable around 0.45, indicating public optimism in the reform's long-term benefits despite media skepticism.

5.4 Mid-2018 (Pre-2019 Elections): Fig 4:(d)

From March to September 2018, Twitter sentiment was notably positive at 0.55, driven by BJP's state-level victories and campaign momentum. In contrast, news sentiment was subdued at 0.40, possibly due to emerging controversies. This suggests a scenario where public discourse shaped the narrative, demonstrat-

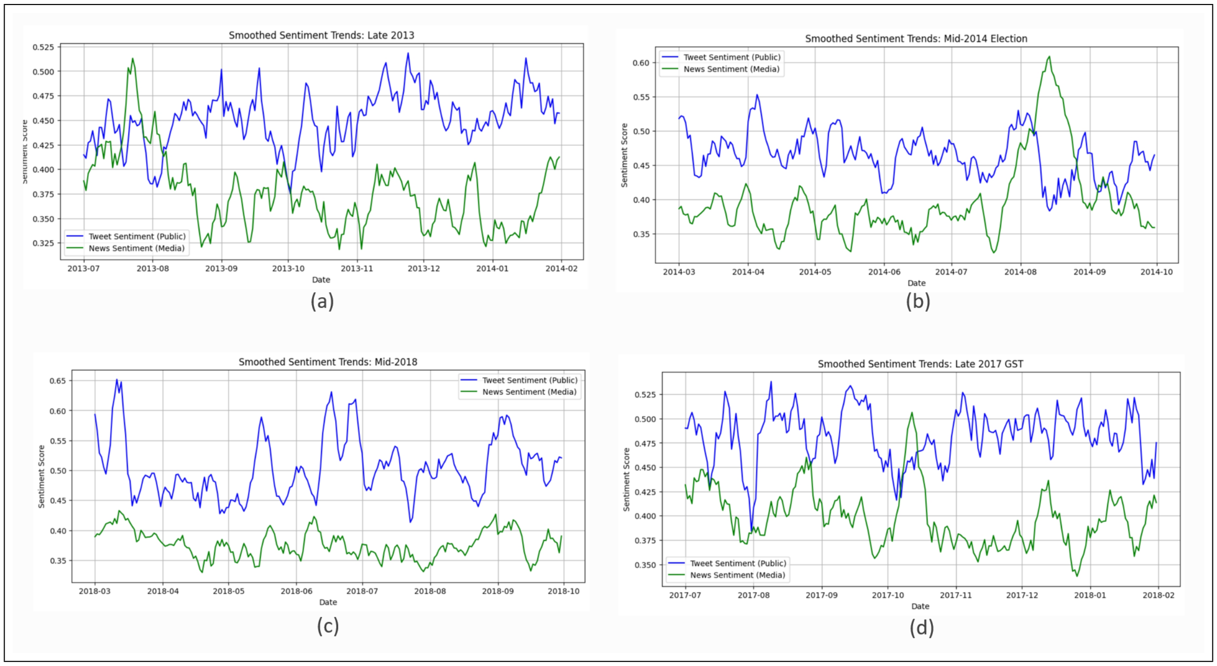

Figure 4: Sentiment trends over key events for BJP (2014–2018).

ing Twitter's influence.

6 Conclusion

This study presents a comprehensive framework for analyzing and forecasting political sentiment toward the Bharatiya Janata Party between 2013 and 2018, leveraging tweets and news articles. Using self-supervised sentiment analysis techniques and advanced models such as prt+ and BiLSTM with BERT embeddings, the framework effectively handled the challenges of unstructured social media data and contextual depth of news content.

The bidirectional analysis between public and media sentiment revealed a dynamic interplay, where traditional media often influenced public opinion, but public sentiment on platforms like Twitter occasionally took the lead. The integration of engagement metrics further enriched the analysis, exposing patterns of sentiment fluctuation, media bias, and the evolving nature of digital political discourse.

Overall, this approach not only enhances our understanding of sentiment evolution over time but also provides actionable insights for political communication. By capturing both public emotions and media narratives, it offers political stakeholders a strategic tool for monitoring sentiment trends, anticipating public reactions, and shaping effective engagement in an increasingly digital political landscape

7 Future Scope and Limitations

7.1 Future Scope

There are several promising directions for extending this work. One key area is enhancing the pRT+ model used for tweet sentiment analysis, which could improve its ability to capture temporal variations and engagement-driven dynamics, providing deeper insights into how public sentiment shifts during major political events. Another avenue involves exploring different model architectures—while a bidirectional LSTM was used in this study, experimenting with unidirectional models or hybrid approaches such as Transformer-based architectures could lead to improved performance and computational efficiency. Additionally, conducting a more detailed time-series analysis of media sentiment could help assess whether the news channel's reporting aligns with public sentiment or exhibits bias, potentially influenced by editorial agendas.

7.2 Limitations

One key limitation of the analysis is the reliance on data from a single news channel, which restricts the ability to capture diverse media narratives across India. Additionally, the study does not incorporate tweet engagement metrics such as retweets and likes, which limits the understanding of how public sentiment spreads and influences discourse. The analysis is also constrained to English-language data, thereby overlooking sentiments expressed in regional languages that are vital to India's multilingual population. Furthermore, the dataset contains temporal gaps that may fail to capture rapid shifts in sentiment during fast-paced political events. Another concern is potential labeling bias, as the sentiment scores are generated using a supervised learning model that may inherit biases from its training data. Lastly, the sentiment trends are interpreted without integrating external political or social events, which could provide essential context to fully understand the observed patterns.

8 Literature review

Over the years, a considerable amount of research has been conducted on sentiment analysis and its evolution over time, espe cially in the context of social media plat forms like Twitter. Researchers have ex plored both supervised and unsupervised methods, utilizing machine learning and deep learning architectures to classify sen timents and understand public opinion. A growing area of interest involves capturing how sentiment changes over time and how such insights can be used for prediction, behavioral analysis, decision-making. and strategic

- ***SentimentArcs:***

A Novel Method for Self-Supervised Sentiment Anal ysis of Time Series The paper "SentimentArcs: A Novel Method for Self-Supervised Sentiment Anal ysis of Time Series" introduces a self-supervised approach for ana lyzing sentiment trends, particu larly in long narratives like nov 2 els. Instead of relying on labeled datasets, the authors create a syn thetic ground truth by aggregat ing outputs from multiple senti ment models and using the ensem ble median as a reference trajec tory. They also propose three evalu ation metrics—Model-Corpus Com patibility (MCC), Ensemble-Corpus Compatibility (ECC), and Model Family Coherence (MFC)—to assess sentiment consistency and model be havior. However, the method is mainly applied to structured fic tional data where sentiment arcs are predictable.

Building on this idea, our project ap plies a similar methodology to polit ical sentiment on Twitter, which is far more dynamic and unstructured. We use multiple pretrained senti ment models to score tweets over a f ive-year period and compare their consistency using MCC-based eval uation. A unique aspect of our ap proach is identifying and removing poorly performing models analysis of MCCscores, leading us to select the most stable model for tracking senti ment. Unlike prior work focused on static classification, our focus shifts toward forecasting sentiment trends, addressing a crucial gap in modeling real-world sentiment evolution over time.

- ***Tracking Sentiment by Time Series Analysis:***

Giachanou and Crestani (2016) explored the appli cation of conventional time series analysis techniques for sentiment tracking on Twitter data. Their work focused on decomposing sen timent time series into trend, sea sonal, and residual components, and introduced additional measures such as sentiment velocity and accelera tion to understand how sentiment changes over time. They also ex amined outlier detection to iden tify significant events responsible for abrupt sentiment shifts. While their study was preliminary and used sim ple lexicon-based sentiment scoring (AFINN), it laid a strong founda tion for modeling sentiment dynam ics and identifying patterns in tem poral sentiment behavior.

In our project, we build on this idea by using pretrained deep learning based sentiment models to generate sentiment scores, rather than rely ing on lexicon-based methods. Al though our primary goal is forecast ing rather than outlier detection, their use of time series decomposi tion techniques (e.g., trend, season ality, moving averages) and senti ment velocity can be applied to im prove signal smoothing and pattern analysis in our temporal sentiment forecasting pipeline. By adopting these ideas, we aim to not only pre dict future sentiment but also bet ter understand the rate and direc tion of sentiment change in political contexts

- ***Sentiment Analysis on Tempo ral Data using Weight Priority Method***:

The paper "Sentiment Analysis on Temporal Data using Weight Priority Method" presents a time-aware approach to senti ment classification by training mul tiple deep learning models—such as Convolutional-LSTM and Stacked LSTM—on data split by time inter vals. The novelty of their method lies in assigning higher weights to 3 models trained on more recent data, using geometric progression to pri oritize recency. This technique acknowledges that user sentiment tends to change over time and that recent data may carry more rele vance when predicting current or near-future sentiment. While their study primarily focuses on sentiment classification using labeled datasets, the concept of time-sensitive weight ing provides a valuable perspective for temporal analysis.

Although our project differs in scope—we focus on forecasting polit ical sentiment using unlabeled Twit ter data—the underlying principle of temporal prioritization is still highly applicable. In our case, instead of training multiple deep models on separate intervals, we aim to ap ply temporal weighting to past senti ment values during forecasting. For example, recent sentiment scores ex tracted from tweets could be given higher weight when fitting time series models like Prophet or LSTM, allowing the model to be more re sponsive to current trends while still accounting for historical patterns. Incorporating this concept can en hance our ability to capture shifts in public mood, especially around fast evolving political events, making our sentiment predictions more adaptive and timely.