Bias Detection in Indian News Media

Team Members

Girish S N (girishsn@iisc.ac.in), Bharath Karanth A (bharathk@iisc.ac.in), Saravanakumar R (saravanakum1@iisc.ac.in), Anmol Gupta (anmolg@iisc.ac.in)

Problem Statement and Motivation

News media outlets are often accused of biased reporting across various categories. With the rapid growth of digital news, readers are exposed to vast and diverse information at their fingertips, often with subtle biases. Biased reporting can influence public opinion and societal harmony.

With increasing polarization and misinformation, unbiased reporting is critical. There is a need for an automated solution that can analyze news articles for bias – offering transparent, multi-dimensional, and interpretable results. This project seeks to provide researchers, fact-checkers, and the public with actionable insights into media bias in Indian journalism.

Objectives

The goal was to achieve a two-fold objective:

- The initial objective was the development of a comprehensive dataset of Indian news articles, complete with rich, multi-dimensional features. Given the absence of any such readily available public dataset, a new dataset was created. Data collection was accomplished via web scraping from various Indian news sources across different languages, with extensive feature engineering performed subsequently.
- The next objective was to detect, identify, categorize, and quantify the bias in Indian news articles across different categories. This was accomplished through various machine learning methods and ensemble techniques.

Workflow

The project workflow involves these stages:

- Data Collection – Scraping article archives into CSV files / MongoDB.
- Data Cleaning – Performing the pre-processing and cleaning of the raw data.
- Translation – Translating non-English articles into English.
- Feature Engineering – Extracting useful features out of the cleaned and translated data.
- Bias Modeling – Assigning bias scores across different categories, through various methods and ensemble techniques.
- Exploratory Visualization – Using the visualization pipeline to generate trends and results through charts.

Please refer to Figure 1 for the Workflow Diagram.

Scraping

As a first step, raw data was collected through the web scraping of online news articles. Multiple Indian news sources were scraped for content; some attempts were successful, while others were not.

The outcome of the process was the creation of a raw dataset, comprising 3,96,739 articles. Please refer to Table 1 for some of the scraping statistics.

News Source	Language	Final Article Count	Approx. Time Taken
The Times of India	**English**	2,023	80 hours *****
The Indian Express	**English**	1,92,050	95 hours
The Economic Times	**English**	1,38,785	80 hours
News18	**English**	49,685	30 hours
Dainik Jagran	**Hindi**	4,268	3 hours
Public TV	**Kannada**	5,059	4 hours
Dinamalar	**Tamil**	4,869	3 hours
Total		**3,96,739**	**295 hours**

Table 1: Scraping Statistics

*Note: Most of the articles scraped for *The Times of India* resulted in failure; hence the time taken is high, but the article count is very low.

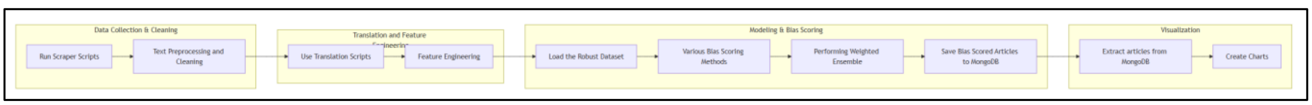

Figure 1: Workflow Diagram

Translation

The non-English articles (Hindi, Kannada, Tamil) were translated to English before further feature engineering and the bias analysis. Various models and APIs were tested, and a summary of the same is presented in Table 2.

Type	Name	Comment	Accuracy	Source
Model	OPUS-MT Marian	Did not support regional Indian languages	N.A.	Open source
API	LibreTranslate	Did not support regional Indian languages	N.A.	Open source
API	Google Translate	Worked very well for all regional languages tested, but **high cost** (About $20 per 1 million characters)	>99%	Proprietary
Model	IndicTrans2	Works well for all regional languages tested	>95%	Hugging Face

Table 2: Models and APIs tried for Translation

The model ultimately used for translation was *ai4bharat/**indictrans2**-indic-en-1B*.

Feature Engineering

In total, 8 feature blocks were added, each with its own number of columns:

- **Temporal** features: Features about when an article appears. (44 columns)
- **Linguistic** features: Low-level properties of the writing style. (35 columns)
- **Keyword-based** features: Counts/weights of specific bias-related words and phrases. (121 columns)
- **Discourse** features: Higher-level structure and framing cues. (30 columns)
- **Semantic** features: Dense representations of the meaning of the article. (2 columns)
- **NER (Named Entity Recognition)** features: Information about which people/places/groups are mentioned. (22 columns)
- **Intersectionality-related** features: Features capturing combinations of identities (e.g. gender + caste + age). (59 columns)
- **Implicit bias** features: Rule-based signals for stereotype-like associations between groups and traits. (98 columns)

All these feature matrices were merged into a single feature table, successively adding the listed number of columns from each block. In total, a total of 411 columns were obtained.

Bias Categories and the Lexicon Used

The project analyzes 10 bias dimensions using keyword-based detection. Figure 2 shows the categories and the interlinks between them.

Further, a curated lexicon of bias-related keywords was created across all these categories. This lexicon list is used for keyword-based features and interpretability. This forms the basis of the calculation of the bias scores.

Bias Scoring Methodology

Figure 3 provides a high-level overview of the bias scoring methods used, and the stage where it comes into the overall picture.

The overall bias score is a weighted average of individual dimension scores obtained after the ensemble:

$$overall\_bias\_score = \Sigma(dimension\_score \times dimension\_weight)$$

Here, $dimension\_score$ is obtained after the ensemble model training, while $dimension\_weight$ is a tunable parameter.

Key Results and Insights

The following are few of the top insights that were generated from the results:

- Political, Gender, and Age Bias Dominate Coverage: These three dimensions consistently appear as the most frequent, highest-scoring, and most persistent bias types across all analyses – from keyword frequency to co-occurrence matrices to temporal trends.

- Coverage is Heavily Skewed Towards Hard News and Sports: Politics and sports dominate article volume, while social issues, environment, health, and education are significantly underrepresented, creating a structural imbalance in topic coverage.
- Implicit Bias Concentrates on Specific Groups with Negative Stereotypes: The most frequent implicit bias patterns target elderly, Muslims, Dalits, and the poor with deficit framing (e.g., "Elderly Weak," "Muslim Violence," "Dalit Poor"), while positive framings are rare.
- Men and Authority Figures Dominate Representation: Male mentions outnumber female mentions by around 3:1, decision-makers appear around 4× more than victims, and authority sources vastly outweigh expert voices – indicating systemic visibility gaps.
- Recency-Bias Exists in the Dataset: Recent years contribute disproportionately more articles, and bias levels have increased since 2013, particularly in regional, socioeconomic, and urban-rural dimensions.
- Sentiment Skews Positive, but Fear Dominates Emotional Framing: While overall sentiment classification leans positive, fear is the most common dominant emotion, followed by joy.
- Religion, Gender, and Socioeconomic Status Have Highest Stereotype Density: When these dimensions are referenced, they are most likely to contain stereotype-related language, even if they don't appear in the highest volume of articles.
- Framing is Polarized, Not Neutral: For specific groups (Muslims, women, poor, Dalits), articles cluster around clearly positive or clearly negative framing scores – neutral coverage is rare, and the poor face the most negative framing.

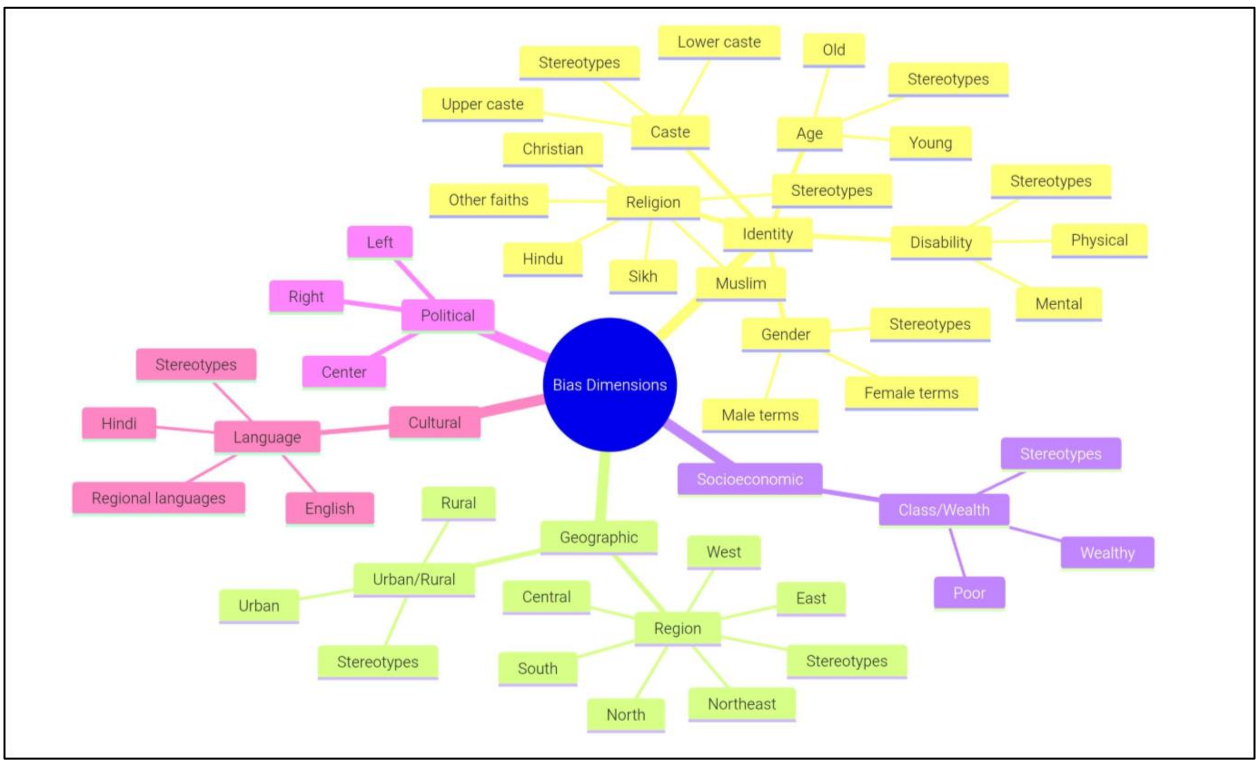

Figure 2: Bias Dimension Categories

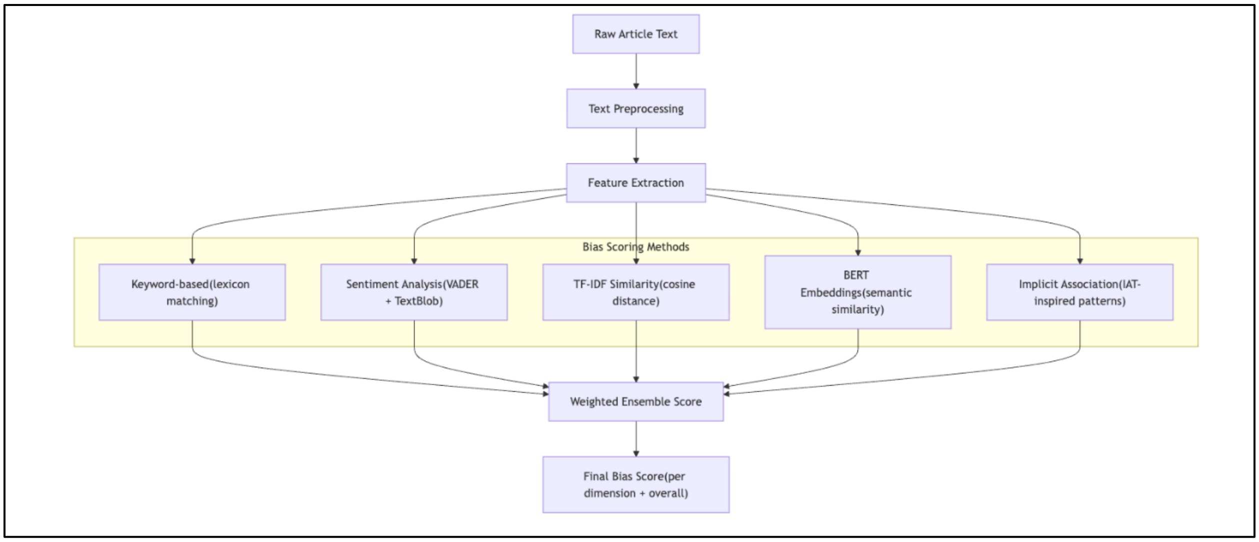

Figure 3: Bias Scoring Methods

Challenges

Some of the major challenges faced during the work done on this project were as follows:

- Identifying scrape-friendly news sites.
- Dealing with IP blocking.
- Improving the scraping performance.
- Ensuring coverage of regional languages.
- Translation of the articles in other languages to English.
- Choosing between native keywords vs translation-first.
- Selecting a suitable language model.
- Deepening NLP/feature engineering skills.
- Constructing a comprehensive India-specific bias lexicon.

Future Roadmap

The future roadmap involves the following:

- Source Expansion: More Indian news outlets and support for more regional languages.
- Increase Bias Categories: Add new numeric bias score fields and the corresponding categorical labels.
- Richer Dashboard Visualizations: Web-based dashboards with interactive graphs and detailed analytics.
- Further Research: Use more models and techniques, and also retrain models.
- Community Feedback: Incorporate impact measurements with feedback from the users.

Contributions by Each Team Member

Here are the individual contributions of each of the four team members:

- **Girish**: Kannada scraper (*Public TV*), Kannada-to-English translation, data cleaning notebook, calendar event integration, discourse and representation analysis, implicit bias detection model.
- **Bharath**: English scrapers (*The Economic Times*, *The Times of India*, *The Indian Express*, *News18*), temporal features EDA, semantic embeddings, ensemble topic classification, intersectionality analysis.
- **Saravanakumar**: Tamil scraper (*Dinamalar*), Tamil-to-English translation, data cleaning scripts, linguistic features EDA, topic-bias correlation analysis, event data generation script.
- **Anmol**: Hindi scrapers (*Dainik Jagran*), Hindi-to-English translation, keyword-based features, NER and entity features, ensemble bias detection model, temporal trends insights.

References

1. Implicit Bias in Religion (Slideshare, 2019)
2. Religion insulates ingroup evaluations (social psychology literature)
3. Measuring gender attitudes: Developing and testing Implicit Association Tests for adolescents in India (PLOS One, 2022)
4. IndiBias: A Benchmark Dataset to Measure Social Biases in Language Models for Indian Context (arXiv:2403.20147)
5. RSF World Press Freedom Index 2025 (global report)
6. RSF – India Country Profile
7. The Habits of Online Newspaper Readers in India (Tewari, 2015)
8. Faux Hate: unravelling the web of fake narratives in spreading hatred and misinformation (Springer, 2024)