Towards Detecting Political Bias in Hindi News Articles

Samyak Agrawal **Kshitij Gupta*** **Devansh Gautam*** **Radhika Mamidi**
International Institute of Information Technology Hyderabad
samyak.agrawal@research.iiit.ac.in
{kshitij.gupta, devansh.gautam}@research.iiit.ac.in,
radhika.mamidi@iiit.ac.in

Abstract

Political propaganda in recent times has been amplified by media news portals through biased reporting, creating untruthful narratives on serious issues causing misinformed public opinions with interests of siding and helping a particular political party. This issue proposes a challenging NLP task of detecting political bias in news articles. We propose a transformer-based transfer learning method to fine-tune the pre-trained network on our data for this bias detection. As the required dataset for this particular task was not available, we created our dataset comprising 1388 Hindi news articles and their headlines from various Hindi news media outlets. We marked them on whether they are biased towards, against, or neutral to BJP, a political party, and the current ruling party at the centre in India.

1 Introduction

Biased news reporting is a widespread phenomenon present in most of the news circulating today .Bias is detected manually, but that is a tedious and time-consuming task; therefore, automation of bias detection in media articles can prove helpful in verifying these articles for their validity more efficiently.

Hindi is an Indo-Aryan language spoken mainly in North India. According to Ethnologue list[1] of most spoken languages worldwide, Hindi ranks third, and a total of 600.5 million Hindi speakers exist in the world[2]. It is also the most spoken language in India with a total of 528,3 million native speakers, which makes up around 43.6 per cent of India's population according to the 2011 census of India[3].

We can observe political bias in news media articles by looking at different factors. We observe biases when the author of the article uses strong language trying to sensationalise an event, is partial to a particular political party, does not give a thorough review of held events etc. The headline in such an article is also essential as it is often filled with bias and is the first thing that catches a reader's attention before they start to read the article. As there is no such dataset annotated for political bias Hindi language, we created our dataset by collecting articles and their headlines from different Hindi news websites. We then annotated the dataset according to whether the article was biased towards or against Bhartiya Janta Party (BJP, the current ruling party at the centre in India) or was neutral.

We present several baseline Machine learning and Deep Learning approaches to detecting political bias on our dataset. We observe that XLM-RoBERTa (Conneau et al. (2020)), a transformer-based model, outperforms other baseline models and achieves a score of 83% accuracy, 76.4% F1-macro, and 72.1% MCC.

The main contributions of our work are as follows:

- We present an annotated dataset consisting of Hindi news articles for political bias detection.
- We propose several baselines using machine learning and deep learning approaches.
- We achieve an F1-macro score of 76.4% by fine-tuning XLM-RoBERTa, a multilingual transformer-based model, on the given dataset.

The rest of the paper is organized as follows. We discuss prior work related to bias detection. We describe the proposed dataset and analyse the annotations. We describe our baseline models and compare the performance of our approaches. We discuss the societal impacts of bias detection. We

*The authors have contributed equally.

[1] https://www.ethnologue.com/guides/ethnologue200

[2] https://www.ethnologue.com/language/hin

[3] https://censusindia.gov.in/2011Census/Language-2011/Statement-4.pdf

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Student Research Workshop, pages 239 - 244

conclude with a direction for future work and highlight our main findings.

2 Related Work

Detection of bias has been studied before with attempts in detecting media bias and its effects on public perception of news and its impact on sociopolitical events like elections.

Misra and Basak (2016) developed an LSTM network model and used it to detect implicit political bias even in the absence of words that relates to either liberal or conservative ideology on two datasets - The Ideological Books Corpus (IBC) and ontheissues (OTI).

LIM et al. (2020) introduces a news bias dataset with sentence-level bias, which allows the development of approaches of bias detection on articles that have subtle bias.

Wei (2020) introduces a dataset of 200,00 sentences regarding Donald Trump and used GloVe vector embeddings to train CNN and RNN to predict the news source of the sentence. They analyze the top 5-grams with their model to gain meaningful insight into Trump's portrayal by different media sources.

Gangula et al. (2019) created a dataset of news articles and headlines collected from Telugu newspapers for bias detection and annotated them for bias towards a particular political party. They also propose a headline attention network model for the detection of bias on their dataset.

Pant et al. (2020) Worked on detecting subjective bias in Wiki Neutrality Corpus (WNC). They propose BERT-based ensemble models for bias detection, which utilizes predictions from multiple models to get better accuracy results.

Some independent organizations also work to fight misinformation. Alt News[4] is a fact-checking website that works to debunk misinformation and disinformation on mainstream social media platforms. Vishwas News[5] is another fact-checking website that is certified by International Fact-Checking Network (IFCN).

3 Dataset Description

We have looked at two major types of biases present in an article while annotating: Coverage / visibility bias and tonality / statement bias (D'Alessio and Allen, 2006). We have annotated our dataset using these biases into 3 categories, biased towards the BJP, biased against the BJP, and neutral if these biases are not visible in the article.

3.1 Target Classes

Coverage bias is concerned with the amount of coverage each side receives over an issue. Articles would at times present only one side of an argument and give undue amount of coverage to that side over the other in order to make viewers side with a particular party. Tonality bias measures the evaluation of a particular actor in the media coverage. In an article, a politician can either be framed positively or negatively changing perception of the general public about them.[6]

3.2 Data Collection

We collected hindi news articles along with their headlines from Indian news websites. We collected these articles from websites of four different news sources. The Wire, The Quint, OPIndia and The Frustrated Indian. The former two are known to be critical of the current government and more liberal media houses, while the latter are known for their pro BJP articles and being more right-wing. We did it to ensure a balance in the number of biased articles for and against the BJP. We collected the links to articles from TheWire using tweets from their Twitter handle @thewirehindi. We used the advanced search feature of Twitter and used hashtags based on the news that was relevant during data collection; for example, #modi, #yogi, #CAA, #BJP, #NRC, #covid etc. For the other three media houses, articles were selected directly from their websites using words relevant to the BJP like modi, bjp, yogi etc. Articles were then scraped from the websites using Selenium[7]. We collected over 8000 articles from all four media websites. Out of these articles, we manually removed irrelevant articles. In the end, a total of 1388 articles were left, which we then annotated for bias.

3.3 Data Annotation

Two annotators did the annotations. Both the annotators are native Hindi speakers and have a good grasp and proficiency in the language. One of the annotators is a self-reported liberal and the other one is a self-reported conservative. Both the annotators were politically up to date with the current

[4] https://www.altnews.in/

[5] https://www.vishvasnews.com/english/

[6] Examples of these biases in our dataset is given in the appendix

[7] https://www.selenium.dev/

	Neutral	For	Against
Articles	234	593	561
Avg #words in headline	15.5	17.8	14.9
Avg #words in article	844	750.9	1222.5
Avg #sentences in article	37.5	31.5	53.4

Table 1: Dataset statistics

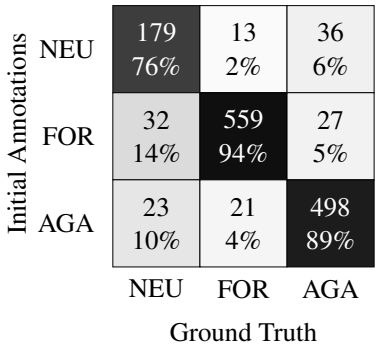

Figure 1: Confusion matrix of the classes annotated by both the annotators. The percentages show the ratio of the ground truth class, which was initially annotated as that class. NEU: Neutral, FOR: For BJP, AGA: Against BJP.

affairs of the country. In the annotation process, we provided both the headline and the article to the annotators. We asked them to read and annotate whether the article and the headline are biased towards the BJP, against it or neutral. We also asked the annotators to do the annotation keeping in mind whether the article exhibits coverage or tonality bias and not deciding on it based on whether the coverage or review is negative or positive. The observed kappa score of the annotations was 0.65. Cases where the two annotators disagreed, were then resolved by a third annotator.

Further, since articles from the same news outlet might have similar biases, we hide the information about the news source and shuffle all the articles before annotating.

3.4 Dataset Analysis

The Dataset contains of 1,388 articles along with their headlines. The general statistics of the dataset are demonstrated in Table 1.

To gauge the difficulty level of each class in the dataset, we analyse the confusion for each class between the two annotators. The results are demonstrated in Figure 1. The confusion matrix indicates that the neutral class is the hardest to detect compared to the other classes. A plausible explanation is the subjective nature of the class, and an article might seem biased to a person while unbiased to another.

4 System Description

In this section, we describe the data splits we use, the evaluation metrics we consider, and the baselines we propose.

4.1 Dataset Splits

The dataset consists of 1,388 articles. We divide the dataset into train and validation sets in the ratio of 10:1 by randomly choosing articles from the dataset. We use the same dataset split for all our models and report the performances on the validation set.

4.2 Evaluation Metrics

We use the following metrics, which are popularly known for classification tasks.

Accuracy is one of the most popular and easy-to-understand metrics. It is a good choice for classification tasks when the data does not suffer from class imbalance.

F1-Score represents a more balanced view, but it could still produce a biased result since it does not consider true negatives. Nonetheless, F1-macro can also handle class imbalance as it gives equal weight to all the classes.

MCC Matthews Correlation Coefficient (Matthews, 1975) takes all parameters of the confusion matrix into account and is less vulnerable to bias. It reports a number in the range -1 to 1, and a key advantage of it is its easy interpretability.

4.3 Baselines

In this section, we provide an overview of the baselines we propose. We experiment with pre-trained language models such as BERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020) on our dataset. We also experiment with traditional machine learning approaches such as SVM, Random Forest to provide exhaustive baselines on the dataset.

4.3.1 MBERT

mBERT is the multilingual version of BERT, which has been trained on a multilingual corpus of 104 languages (including Hindi) using articles from Wikipedia as its training corpus. We leverage the

Model	Accuracy	F1-macro	MCC
mBERT	80.2 $\pm$1.4	72.4 $\pm$2.1	67.1 $\pm$2.2
XLM-RoBERTa	**83** $\pm$**1.1**	**76.4** $\pm$**1.3**	**72.1** $\pm$**1.8**
XLM-RoBERTa (Hindi)	79.2 $\pm$1.5	72.5 $\pm$3.1	65.8 $\pm$2.6
IndicBERT	78.9 $\pm$1.2	69.2 $\pm$4.5	65.5 $\pm$2.1
SVM	78.7	59.6	64.6
Logistic Regression	77.1	55.1	61.5
Random Forest	78.7	59.6	64.6

Table 2: Mean and std dev are reported across five runs of all the models.

Hindi pre-training of the model and fine-tune the model on our dataset.

We use a `[SEP]` token between the headline and the contents of the article to prepare the input for the transformer network. For classification, we attach a feed-forward network on the `[CLS]` token embedding with two linear layers having the model's default dropout of 0.1 and *Tanh* activation layer in between. To train our model, we use Adam optimizer with a learning rate of $1e^{-5}$ and a batch size of 16 with a maximum sequence length of 256. We use the standard cross entropy loss to train our model.

4.3.2 XLM-RoBERTa

XLM-RoBERTa (Conneau et al., 2020) is the multilingual version of RoBERTa (Liu et al., 2019) which is an optimized version of BERT. XLM-RoBERTa has been pre-trained on 2.5TB of filtered CommonCrawl data containing 100 different languages. We leverage the Hindi pre-training of the model and fine-tune the model on our dataset for bias detection.

Since multilingual versions often perform slightly worse than their monolingual counterparts, we also experiment with a monolingual version of XLM-RoBERTa (Jain et al., 2020). The model has been pre-trained on 3GB of Hindi monolingual data majorly taken from OSCAR (Ortiz Suárez et al., 2020).

To train the models, we use the same classification network and training parameters as mentioned in Section 4.3.1.

4.3.3 IndicBERT

IndicBERT (Kakwani et al., 2020) is a multilingual model based on ALBERT (Lan et al., 2020) which has been pre-trained on 12 major Indian languages. The model has much fewer parameters than mBERT and XLM-RoBERTa, but it can still achieve similar performances or even better in most of the tasks.

To train the model, we use the same classification network and training parameters as mentioned in Section 4.3.1.

4.3.4 SVM

Support Vector Machines are models for classification and regression problems. First, the textual data is transformed to a set of features by using methods like Bag of words, Bag-of-n-grams, or Tf-Idf. Later, the classification model is applied on the transformed features. The kernel we used is the Radial Basis Function (RBF) kernel which is a non-linear kernel. The RBF kernel function computes the similarity between two points ($\mathbf{x}$, $\mathbf{x}'$) or how close they are to each other. This kernel can be explained as:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (1)$$

where γ is a free parameter.

We first generate the count matrix of all the tokens in the text. We use Term frequency (TF)-Inverse Document Frequency (IDF) to normalize the count matrix and use it to train our model.The regularization parameter is set to 1. The loss function used is hinge loss.

4.3.5 Logistic Regression

Logistic Regression is a another supervised learning approach like SVM which differs by using the weighted combination of the input features and passes them through a sigmoid function.

Similar to SVM, here we use TF-IDF to get features from the articles which are then given to our model. We use the standard cross entropy loss to train our model.

4.3.6 Random Forest

Random forests is a classification algorithm which creates an ensemble of decision trees. It uses bag-

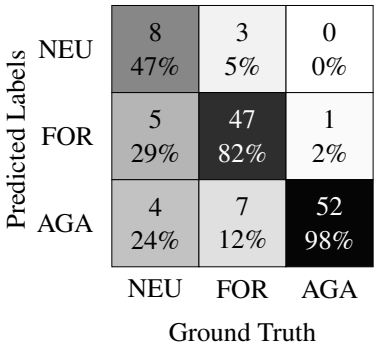

Figure 2: Confusion matrix of the classes predicted by the best performing model in the validation set. The percentages show the ratio of the target class, which was predicted as that class. NEU: Neutral, FOR: For BJP, AGA: Against BJP.

ging and feature randomness to build each individual tree and then use the predictions of the forest of trees which is more accurate than the prediction of any individual tree.

Similar to SVM, here we again use TF-IDF to get features from the articles which are then given to our model. We use the standard cross entropy loss to train our model. The quality of the split is measures using the gini criterion. The minimum sample split was kept at 2.

5 Results and Discussion

We report the results of our models in Table 2. We observe that the deep learning models perform better than the machine learning approaches. The results further indicate that even simpler models can give decent performances on the given problem.

To further analyse the results, we compare the class-wise results of the best models. We show the confusion matrix of the predictions compared to the ground truth values in Figure 2. The model is performing very well on the biased classes but suffers heavily on the neutral class. We observe the same pattern during the annotation process and believe that predicting whether an article is unbiased is comparatively more challenging than predicting the type of bias.

5.1 Societal Implications and Limitations

Online news in today's day and age strongly influences the general public's opinion. Ideally, news media should report the news objectively and from a neutral standpoint, but that is seldom the case. The news these days is highly subjective, biased and thus, these media companies put in a lot of opinionated information through sections of society. Biased news can have long-term and far-reaching implications for public opinion on societal issues and how they view government policies, laws and elections (Baum and Gussin, 2005; Bernhardt et al., 2008). People should have access to an unbiased and objective form of news reporting. In India, we can see news channels and news websites online pushing out one-sided and highly opinionated news. Chadha et al. (2019) discuss the discourse of several news portals with an inherent bias towards right-wing politics and how they talk about their "aims to provide a counter to the mainstream media narrative about India" which they consider to be "left-liberal" and "pseudo-secular". They also discuss how the members of political parties fund these websites to carry out propaganda on their behalf. This shows that instead of news sources providing unbiased news, we have news portals at two opposite sides of the political spectrum which will publish information and make opinion pieces keeping their political leaning in mind. Such biased news portals make the detection of political bias even more important and relevant in today's times.

Our system is trained on articles from a limited number of sources and thus might not be fitted well to make predictions on news articles from other sources. Also, predictions from our model which might be incorrect can be used to accuse certain media houses as biased. Thus, our system should rather be used as a method to filter out potentially biased articles from a larger set of articles rather than using it as a gold standard to mark articles as being biased.

6 Conclusion

In this paper, we proposed a dataset to detect biases in Hindi news articles. We analysed the difficulty level of each class, and our experiments indicate that detecting whether an article is unbiased is a more challenging problem than detecting the type of bias. Further, we provided several baseline models on the proposed dataset and found out that multilingual deep learning models outperform other approaches by a large margin and should be the choice for performance metrics. We perform error analysis on the best performing model to further understand the shortcomings of our proposed system. Lastly, we also discussed the ethical and societal implications of the proposed work.

As a part of future work, we aim to extend the

system by shifting our focus from a particular political party and propose a general approach for any set of political parties.

References

Matthew A Baum and Phil Gussin. 2005. Issue bias: How issue coverage and media bias affect voter perceptions of elections. In *Meeting of the American Political Science Association. Washington, DC: Apsa*. Citeseer.

Dan Bernhardt, Stefan Krasa, and Mattias Polborn. 2008. Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5-6):1092–1104.

K Chadha, P Bhat, and Shakuntala Rao. 2019. The media are biased: Exploring online right-wing responses to mainstream news in india. *Indian journalism in a New Era: Changes, challenges and perspectives*, pages 115–139.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Dave D'Alessio and Mike Allen. 2006. Media Bias in Presidential Elections: A Meta-Analysis. *Journal of Communication*, 50(4):133–156.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Rama Rohit Reddy Gangula, Suma Reddy Duggenpudi, and Radhika Mamidi. 2019. Detecting political bias in news articles using headline attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 77–84, Florence, Italy. Association for Computational Linguistics.

Kushal Jain, Adwait Deshpande, Kumar Shridhar, Felix Laumann, and Ayushman Dash. 2020. Indic-transformers: An analysis of transformer language models for indian languages. *CoRR*, abs/2011.02323.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online. Association for Computational Linguistics.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. Albert: A lite bert for self-supervised learning of language representations. In *International Conference on Learning Representations*.

Sora LIM, Adam JATOWT, and Masatoshi YOSHIKAWA. 2020. Creating a dataset for fine-grained bias detection in news articles. In *Forum on Data Engineering and Information Management*, volume 12, pages 1–35.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

B.W. Matthews. 1975. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2):442–451.

Arkajyoti Misra and Sanjib Basak. 2016. Political bias analysis.

Pedro Javier Ortiz Suárez, Laurent Romary, and Benoît Sagot. 2020. A monolingual approach to contextualized word embeddings for mid-resource languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1703–1714, Online. Association for Computational Linguistics.

Kartikey Pant, Tanvi Dadu, and Radhika Mamidi. 2020. Towards detection of subjective bias using contextualized word embeddings. In *Companion Proceedings of the Web Conference 2020*, WWW '20, page 75–76, New York, NY, USA. Association for Computing Machinery.

Jerry Wei. 2020. Newb: 200, 000+ sentences for political bias detection. *CoRR*, abs/2006.03051.