Tackling Misinformation: Studying COVID-19 Narrative by Indian News Media using Headline Stance Detection as a first step

Tazbia Fatima

Columbia University

Abstract

Journalism's first obligation is to the truth[1] and its loyalty lies to the citizens. According to the American Press Institute website, "the purpose and importance of journalism is to inform society with the information they need to live their lives. Information that they need to make decisions about different things. The outbreak of the Novel Corona Virus put more onus on delivering truthful journalism to assist the public in taking crucial decisions that impact both their lives and the larger society. However, there has been dissemination of misinformation or ambiguity by news articles and agencies. COVID-19 coverage has dominated the frontpage ever since and in 2021, overall media consumption among U.S. adults is estimated to be around 11.1 hours a day — a 4.35**%** increase from 2019.[5]

However, roughly six in 10 news consumers have acknowledged that they are only headline-gazers.[2] Thus, increasing the importance of having clear headlines that convey the truth.

In the context of COVID-19 news coverage by the Indian Media, this paper studies Stance Detection, the technique of identifying the relationship between a headline and news articles to determine the credibility of media for an average individual who skims through the news. It explores the validity of a pre-COVID-19 trained classifier in predicting the stance of COVID-19 specific articles. As the detection and prevention of fake news of COVID-19 presents specific challenges, our conclusion identified potential challenges and research directions.

Keywords: Stance Detection, Misinformation, Classification, COVID-19, Feature Engineering

1. Introduction:

Fake News and misinformation have been a prominent part of research regarding media's credibility. There are various misinformation classifiers and disinformation detectors built in the past to combat, alleviate and study this problem. About three-quarters of Americans who say they follow news and current events agree that fake news is a big problem today, according to findings from Deloitte's recent Digital Media Trends study.[3]

While the impacts of fake news are noticed in large scales during major events like Presidential Elections, it has become all the more critical in its direct impact to individuals since the COVID-19 outbreak. There has been a global spread of false information and conspiracy theories about Coronavirus, with World Health Organization terming this phenomenon an 'infodemic', or 'an overabundance of information— some accurate and some not— that makes it hard for people to find trustworthy sources and reliable guidance when they need it.'[4]

In India, the country with the second highest number of COVID-19 deaths, second to United States of America, the pandemic created an atmosphere of panic and mismanagement. The high rates of vaccine reluctancy are also linked to the widespread fear mongering.[7]

A report on 'Localized Misinformation in a Global Pandemic' by The Empirical Studies of Conflict Project (ESOC) at the Woodrow Wilson School of Public and International Affairs at Princeton University states that "Misinformation in Southeast Asia primarily responded to regional politics. False narratives played on the existing religious and ethnic diversity in the region, while frequently adopting racial and anti-immigrant undertones as well." [8]

In April, Islamophobic misinformation took root in the country, accusing Muslims of #CoronaJihad – a term used to imply terrorism. This hashtag began trending on Twitter and was covered extensively by the Indian journalists after several members of a Sunni religious movement, Tablighi Jamaat, tested positive for COVID-19 after a gathering held in New Delhi. Soon after, Academics reported on this incident saying "at the outset of the pandemic, extant hatred, suspicion, and misinformation directed at Indian Muslims easily incorporated lies about COVID-19 (Yadav et al., 2020). [9]

Individuals, politicians, and media outlets across the region were found guilty of propagating untested, unfounded and bizarre medicinal cures like homeopathic drug coronil, animal excreta and homemade spice teas. Doctors in India had to warn against the practice of using cow dung in the belief that it will ward off COVID-19, saying there is no scientific evidence for its effectiveness and that it risks spreading other diseases. [10]

In India, the Government of India's Press Information Bureau regularly debunks coronavirus-related misinformation on its Twitter account, but this seems to be the government's only method to systematically expose false information.

Fake news can hurt the individual and the community and headlines shape the way a reader defines the information presented in an article.[11] Therefore, in light of these consequences, as well as concerns that the effect of misinformation on reducing public trust is particularly dangerous in the midst of a global health crisis, studying the accuracy of headlines to the actual news article provides valuable insight.

1.1 Definition of Fake News:

Fake News is a broad spectrum with multiple variants identified[12] (Tandoc et al., 2018; Rubin et al., 2015). as:

1. Clickbait: Headlines that easily capture user attention without fulfilling user expectations. They often use slang and question-based titles. Their main aim is to increase the reader or engagement count.

2. Propaganda: Deliberately biased information designed to mislead the audience and promote propaganda. (Rubin et al., 2015).

3. Satire or Parody: Exaggerated Information published by several websites for the entertainment of users such as "The Onion" website.

4. Misleading Headings: Ambiguous headlines with sensationalist or misleading information.

5. Slanted or Biased News: Information that describes one side of a story by suppressing evidence that supports the other side or argument.

1.2 Related Work:

Existing work related to misinformation detection and classification can be further classified into the following four paradigms:

1. Hybrid Approach: [13] This approach uses both human and ML approaches for the detection of fake news.

2. Feature-Based: This approach employs three sub-categories – account-based, context and content-based and Text categorization after feature engineering of the data.

3. Network Propagation: This approach includes potential methods for discovering, flagging and stopping the propagation of fake news in its infancy.

4. Knowledge-Based: This approach uses human expert knowledge to supplementing AI models for decision-making.

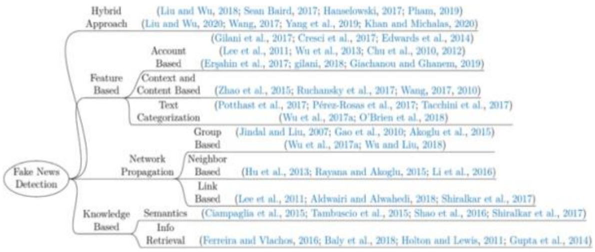


Figure 1: Misinformation classification paradigms[14]

For the variant of misleading headlines, a study proposed co-training approaches for a problem of detecting ambiguous headlines from the pair of title and body text (Wei and Wan 2017). However, the researchers utilized a small set of news articles that are manually labeled and not shared for the public.

Previous work in this field have also proposed classification of headline-article relations as incongruity detection. One notable work is "Detecting Incongruity Between News Headline and Body Text via a Deep Hierarchical Encoder"[16] (Yoon et all 2018). It inspects the textual relationship between a headline and each paragraph respectively, rather than examining the relationship between the headline and whole article content at once. It implanted unrelated or topically-inconsistent content into body text of original news articles making headlines distinct stories with its article content. Incongruency is a characteristic of clickbait as it essentially checks relationship between only mismatched headline-article pairs.

In this paper we focus on misleading headlines and the feature-based context and content-based text-categorization paradigm to identify relationship of a headline to its correct article.

TALOS[17] -The first ranked team in the Fake News Challenge-1[18] conducted by Cisco in 2017. Their approach to create a stance detection system is based on a 50/50 weighted average ensemble combining a gradient-boosted decision tree model fed with text-based features from the headline and the body pair, and a deep learning model based on one dimensional Convolutional Neural Network (CNN) with Google News pre-trained vectors.

In this paper, we have chosen to implement the classifier built by Talos to classify COVID-19 articles covered by Indian journalists. We also aim to identify if the classifier is robust enough to predict the stance of COVID-19 related news articles even though it was built with a pre-COVID vocabulary and hence did not have the common scientific terms used in COVID-19 news coverage.

1.3 Methodology:

Stance Detection is about predicting the relation of a headline towards the paired news article (body of the article). Dean Pomerleau, one of the organizers of the Cisco's Fake News Challenge, explained in a Mediashift interview that "[...] the goal [of stance detection] is to determine which[headline] has the best argument, not just which is the most popular or widely cited or read, the way a search engine does."

The idea of the system is to classify the headline by labeling the relationship an article body has to its headline/claim. Thus, the four classes are defined as:

1. Agrees: The body text agrees with the headline.
2. Disagrees: The body text disagrees with the headline.
3. Discusses: The body text discusses the same topic as the headline, but does not take a position
4. Unrelated: The body text discusses a different topic than the headline. (Similar to incongruency)

In this research we implemented the Gradient-Boosted Decision Trees (GBDT) model of TALOS, as Deep neural networks need more amount of data to show their relevance.

We input headlines and full body of an article to the model. They are then converted to unigrams, bigrams and trigrams before text-based features are derived from both.

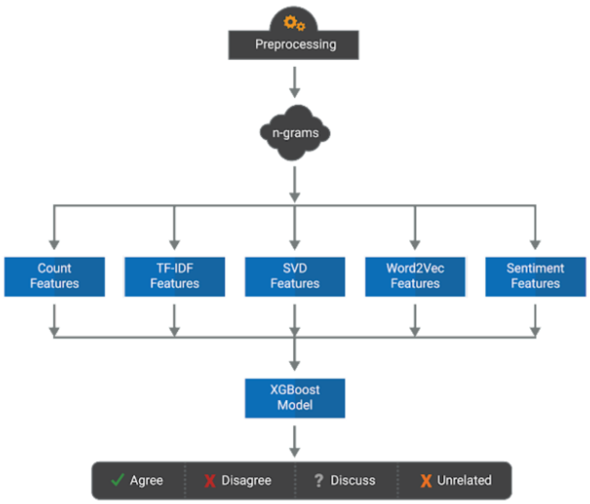


Figure 2: Gradient Tree Boosted Model Methodology sourced from TALOS

The features extracted from the texts are as follows:

1. Count Features: It counts how many times a gram appears in the headline, how many unique grams there are in the headline, and the ratio between the two. Example: ['count_of_Headline_unigram', 'ratio_of_Headline_bigram_in_articleBody', 'count_of_Headline_trigram_in_articleBody', etc.]

2. TF-IDF Features: Calculating the Term-Frequency of each gram and normalize it by its Inverse-Document Frequency. Cosine similarity between the headline vector and the body vector is calculated.

3. SVD Features: Finding the latent topics involved in the corpus and represent each headline/body text as a mixture of these topics. The cosine similarities between the SVD features of headline and body text are also computed.

4. Word2Vec features: Used to find synonyms by grouping mathematically similar word vectors together. It starts by creating distributed numerical representations of word features.

5. Sentiment Features: Uses the Sentiment Analyzer in the NLTK package to assign a sentiment polarity score to the headline and body separately. For example, negative score means the text shows a negative opinion of something. Example: ['fake_exist', 'fraud_exist','deny_exist', 'denies_exist','doubt_exist', 'pranks_exist']

These features are then fed into Gradient Boosted Trees to predict the relation between the headline and the body.

1.4 Implementation:

1.4.1 Data Collection:

Seven hundred and seventy-seven headlines and their corresponding full-text news article were retrieved via NewsApi. The articles were filtered by COVID specific keywords such as 'COVID-19', 'Coronavirus', 'Covishield', 'Covaxin', 'Pfizer', 'Astra-Zeneca', 'Omicron' and 'Delta'. The articles in the sample set were published between November 10, 2021 to December 11, 2021. They belong to the news agencies Times of India, Press Trust of India, Indo-Asian News and a few other news agencies wired through the Times of India.

The dataset was split into a train-test set by the ratio 75% and 25% respectively, with 579 headlines-body pairs and 198 headline-body pairs in the train and test set respectively.

The train set headlines were manually labeled into one of the four classes. The class distribution was slightly more spread out to account for the articles that had straightforward, factual coverage of COVID-19 such as No. of Cases or Vaccine Efficacy.

Training Set	Class Distribution			
	Agree	**Disagree**	**Discuss**	**Unrelated**
FNC-1 original train set	7.36%	1.68%	17.83%	73.13%
New COVID-19 train set	48.87%	3.97%	11.91%	35.23%

Table 1: Comparison of Class Distribution of Original pre-COVID Training set and New COVID-19 Training set

1.4.2 Execution:

<u>1.4.2.1 First Iteration:</u>

The model was trained on the original training set of the Talos team consisting of 49,972 headline-article pairs as training data. The training dataset was a match between 1648 unique headlines and 1683 unique article bodies. The model pre-processed the training set, extracted the features and predicted the class of the 198 headline-article pairs of the test set.

However, on manual cross-validation of the set, the articles labeled as “discuss” were actually unrelated between the headline and article body. This was because the headline-article pair had multiple words or topics that were common to both vectors such as the words ‘COVID’, ‘vaccine’ or ‘efficacy’ but their context was different. For example, the headline “*Coronavirus vaccination: Suctioning COVID vaccines may provide better response than needles, finds study*” was classified as “discuss” corresponding to the body “*People who receive the Pfizer or AstraZeneca COVID-19 vaccine have antibody levels significantly higher than those infected with the SARS-CoV-2 virus, according to a study published in Scientific Reports journal on Monday…*”

This is a nuance of the COVID data, that has limited scientific vocabulary.

<u>1.4.2.2 Second Iteration:</u>

Keeping in mind that the classifier was built in 2017, when the COVID pandemic was non-existent and not in the vocabulary of the training set, the model was then trained on the new COVID-19 specific manually-labeled dataset.

In addition to this, a random selection of 100 headlines and 100 articles from this 495 was randomly matched and manually labeled into one of the four classes to account for the similar words between unrelated headline-article pairs. They mostly fell into the “unrelated” class as they were a random match, so considering the above example, the random pairs were labeled as unrelated even if they had common words. The model then predicted the classes of the test set. The test set was cross-validated manually.

1.5 Results and Analysis:

1.5.1 <u>Analysis of Data according to the classification by Model trained on COVID-19 dataset and cross-validation:</u>

To analyze the result, it is essential to understand the distribution of articles for each news agency in the sample set.

From the 198 articles in the test set, The Times of India including The Economic Times and Times News Network has the highest proportion of articles i.e., 117 with 91, 20 and 6 articles respectively. These three publications are under the same news company. Whereas Press Trust of India has the second highest number of articles in the sample, 53. These agencies occupy the bulk of the sample set.

The Times of India	91
Press Trust of India	53
The Economic Times	20
Asian News International	16
Indo-Asian News Service	7
Times News Network	6
The Pioneer	1
YouTube	1

Figure 3: Total number of articles in test-set for each News Agency

It is not robust to analyze the result for “The Pioneer” and “YouTube” as they only have one article in the set and hence, they will be ignored.

Analyzing the results by percentage of articles by class for each author or News agency:

The Times of India, The Economic Times and Times News Network have a mean of 59.43% non-ambiguous and non-misleading headlines. Whereas, Press Trust of India has 62.26% “agree” headlines with no headlines in the “discuss” category. Indo-Asian News Network has 72% in “agree” however it is an analyses of only 7 articles for it.

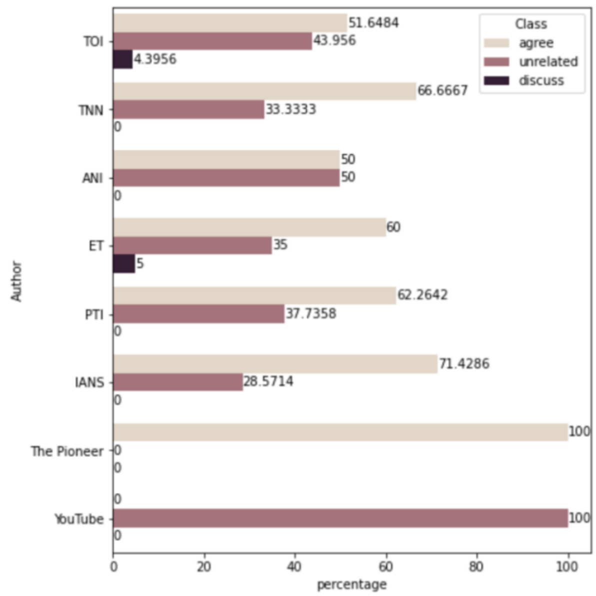


Figure 4: Percentage of articles by class for each News Agency (Author)

In the dataset, the articles were in the categories of Health, Business, Economy, Science and General.
The "General" category has most news based on statements, interviews or opinions of politicians or academicians. Upon cross-validation, they fell under the "agree" category as most headlines restate the first sentence of the body.
The categories Health, Business and Economy had most ambiguity in articles that were explaining the results of a finding or a study. For example: This one sector can perform well in 20 22 was classified as "unrelated" even though it is a "discuss" stance. Thus, showing the ambiguity in the headline, which can potentially be mislead ing

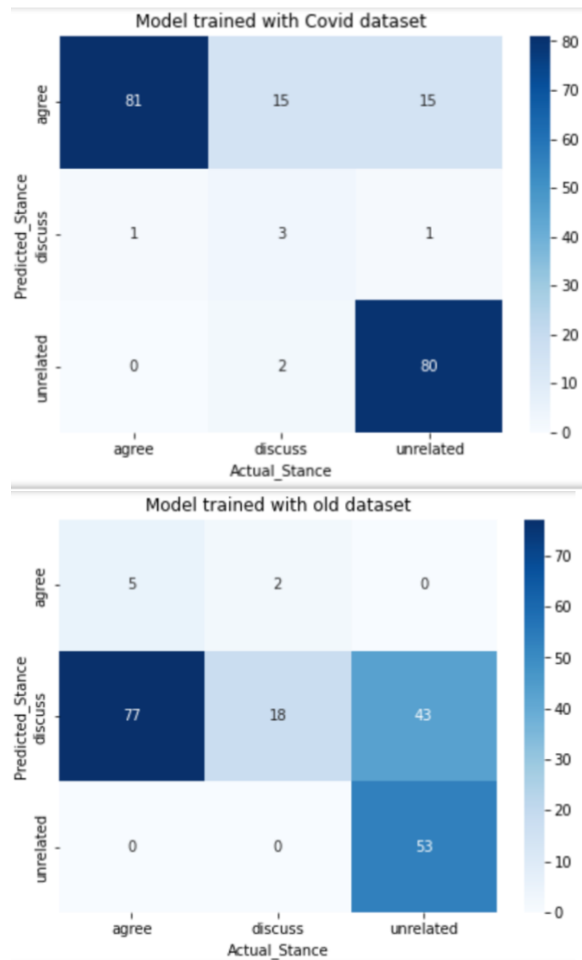
Overall, the results from this sample set lean towards the claim that "The Times of India" was the most trusted news brand in 2021.[14]
Overall mean % of straight-forward headlines in the test set is 56.06%.

The robustness of these claims can be challenged due to certain limitations and caveats of this research explained in the next section.

1.5.2. Analysis of Classifier:

Through our two iterations, we found that the classifier trained on the pre-COVID dataset performed poorer than the same model trained on a much smaller but COVID-19 specific dataset.

Figure 5: Confusion Matrix for both iterations



	Model trained on COVID-19 dataset	Model trained on pre-COVID-19 dataset
Precision	0.768	0.614
Recall	0.657	0.504
Accuracy	0.8282	0.383

Table 2: Overall model performance metrics

Table 3.1.1: Class-wise Performance Metrics for Model Trained on COVID-19 Dataset

	precision	recall	f1-score	support
agree	0.714286	0.060976	0.112360	82.000000
discuss	0.130435	0.900000	0.227848	20.000000
unrelated	1.000000	0.552083	0.711409	96.000000
accuracy	0.383838	0.383838	0.383838	0.383838
macro avg	0.614907	0.504353	0.350539	198.000000
weighted avg	0.793839	0.383838	0.414473	198.000000

Table 2.1.2: Class-wise Performance Metrics for Model Trained on pre-COVID-19 Dataset

From the above confusion matrices, we can see that re-training of the classifier is required to have a more accurate classification.

Another interesting point of this classifier is its definition of the classes:

1. From Manual Cross-Validation, we recognized that the headlines that are exact phrases taken from the body are classified as "agree" which clearly represent a non-ambiguous, non-misleading headline.
2. In real word context when trying to analyze the credibility of news agencies, there will be no headline "unrelated" to its body, however, there may be certain slang words used that are common to the region or people there and these are not recognized as synonyms by the Word2Vec feature extraction or during the lemmatization in the pre-processing of the train set.
   Example: The headline "*Amid virus scare, bulls & bears fight it out on D-Street*" refers to an article talking about Dalal Street in Mumbai, the address of the Bombay Stock Exchange. It was wrongly classified as "unrelated".
3. The class "disagree" checks for exactly negative words such as "debunked", "don't", "despite", "prank" etc. This class is looking for a strong rejection in the headline corresponding to the claim in the article. It was noted, that not even one headline-article pair was classified into the disagree class by the classifier with either of the training set.
4. The most contextually interesting class is the "discuss" class, which contains headlines that discuss only a certain part of the article instead of capturing the main point of the article, or a headline that uses heavy slang. For example: The headline "*COVID on the prowl with no telltale signs this time*" was related to an article about how certain members above the age of forty-five were COVID-19 positive but had no symptoms in a city.

Overall, the classifier's label "agree" is a straight-forward headline that states the gist. It is not misleading because if someone reads the headline only, they'll get an accurate understanding of the article body, and can thus, be binary classified as "not misleading". While, "Discuss" is the class that identifies ambiguity and needs further analysis.

1.6. Limitations:

There are two main limitations to the robustness of this analysis of COVID-19 news coverage by the Indian news media:

1. Limited Data: To make a claim that 56.06% of the COVID-19 articles covered by Indian news are not misleading or "agree" to the body, we need to have substantial amount of data, spread almost equally across, all categories and agencies. For the scope of this research, more than one month's articles could not be collected or labeled due to logistic reasons.

2. Manual labeling: It would be ideal to have multiple journalists label or classify the headline-body pairs and then take majority rule as the class. It would increase the trust in the cross-validation.

1.7. Conclusion and Discussion:

Through this research, there are three points to be noted:

1. Stance detection is a clear and objective way to identify the relation between a headline and its article, however, it does not completely

encapsulate the classification of whether a headline is misleading or not.

2. Stance Detection classifier of Talos requires re-training with jargon specific to the COVID-19 for better classifications.

3. General lemmatization fails to account for regional slang and synonyms.

For, a robust categorization, a proposed future model should be one that analyses a headline-article pair for all variants of fake news or misinformation and computes final classification as misleading or not.

Post Stance Detection, we classified our test set as "clickbait", "not_clickbait", using a Naïve Bayes based Clickbait Detection model. The result shows that most headlines classified as "unrelated" or "discuss" during Stance Detection were also classified as "clickbait". Example: "*COVID on the prowl with no telltale signs this time*" the headlines is classified as "unrelated" and "clickbait"

This recognition, gives a possible research direction of implementing a collection of analysis tools in one platform which helps human fact checkers and normal users in producing better judgement based on multiple aspects.

For example, if a headline is classified as "discuss" or "disagree" to its article body and classified as having "clickbait" text or biased text, then it can be given a score of accuracy, and classified as misleading or disinformation. Its accuracy score will be lower in comparison to a headline classified as only "discuss" and "not clickbait".

Therefore, performing stance detection coupled with detection for multiple variants of fake news such as click-bait detection and biased-news detection might provide a more contextually accurate classification that's closer to the understanding of a human news reader.

References:

1. The Elements of Journalism, Bill Kovach and Tom Rosenstiel identify the essential principles and practices of journalism.
2. https://www.washingtonpost.com/news/the-fix/wp/2014/03/19/americans-read-headlines-and-not-much-else/
3. https://www2.deloitte.com/us/en/insights/industry/technology/study-shows-news-consumers-consider-fake-news-a-big-problem.html/#endnote-2
4. (WHO, 2019).
5. https://www.visualcapitalist.com/how-media-consumption-has-changed-in-2021/)
6. Schild Et al 2020
7. https://www.sciencedirect.com/science/article/pii/S2666354621001782
8. https://esoc.princeton.edu/publications/localized-misinformation-global-pandemic-report-COVID-19-narratives-around-world
9. Yadav et al., 2020
10. https://www.reuters.com/world/india/indian-doctors-warn-against-cow-dung-COVID-cure-2021-05-11/
11. https://link.springer.com/article/10.3758/BF03210784#page-1
12. Tandoc et al., 2018; Rubin et al., 2015
13. https://www.sciencedirect.com/topics/computer-science/machine-learning-approach
14. https://www.sciencedirect.com/science/article/pii/S1084804521001326#sec2.1
15. https://www.statista.com/topics/8332/news-consumption-trends-in-india/#dossierKeyfigures
16. https://arxiv.org/abs/1811.07066
17. https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html
18. https://github.com/FakeNewsChallenge/fnc-1
19.