Array 6 (2020) 100025

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/journals/array/2590-0056/open-access-journal

Media bias detection and bias short term impact assessment

Swati Aggarwal [*,1], Tushar Sinha [1], Yash Kukreti [1], Siddarth Shikhar [1]

Division of Computer Engineering, Netaji Subhas University of Technology, Azad Hind Fauj Marg, Sector-3, Dwarka, New Delhi, 110078, India

ARTICLE INFO

Index Terms:
Media bias
Biased opinion
User conditioning
News impact

ABSTRACT

Today, media outlets are known to report news in a biased way, potentially affecting the beliefs of news consumers and altering their behaviors. Therefore, tracking bias in everyday news and building a platform where people can receive neutral and unequivocal news information is important.

This research investigates media outlets for subjectivity versus objectivity by examining reported news events from their Twitter handles. The study subsequently proceeds to show how such subjective news articles program news consumers and condition their opinions. Eventually, a system which aids the detection of alarming biases in media is proposed through a unique bias short-term impact score calculation mechanism.

With reference to a real-world event, 'Demonetization in India', a significant fraction of news tweets is found to be subjective in nature which are further observed as opinion conditioning agents for their consumers. Alarming biases are then detected in tweets in comparison with a neutral baseline.

1. Introduction

Opinions play a key role in decision-making processes. When a customer chooses a product or service, he is influenced by the views of others. Traditionally, people relied on the experiences of family members and peers in order to make better decisions. In today's era of social media, individuals and groups not only exchange content and engage in discussions, but also offer their points of view on different products, services or people. Websites such as Facebook and Twitter offer a continuous stream of personal opinions on the most varied range of topics. This helps people to learn about experiences of individuals all over the world.

In the present day where social media has become a powerful means for expressing opinions, what is feared is how it is slowly programming users' behaviors. News, among all other means, has grown to become one of the most influential carriers of motivated propaganda.

The burgeoning of fake news is an evident side effect of this phenomenon. The US presidential election campaign in 2016, notably, is one of the numerous incidents which have been plagued with this menace in recent times. According to The Telegraph, fake news is now seen as one of the greatest threats to democracy, free debate and the Western order. News reporting, which functions on the ideals of objectivity, needs to be restored to its principled and lucid past.

[1] Wikipedia defines *media bias* as "the perceived bias of journalists and news producers within mass media in the selection of events and stories that are reported, and how they are covered."

Media has the power to influence opinions of masses which in turn conditions and influences their day-to-day activities. In recent years, a lot of controversy has risen over the credibility of media outlets in reporting news. This calls for a need to detect bias in everyday news and to develop a platform so that people can receive balanced and safe news.

This research relies on Twitter (one of the most prominent social networking medium present today) for its analysis. Tweets by media outlets are indicative of the type of news they report. This study targets some popular media outlets and journalists in India, and analyzes their tweets. It analyzes how biased opinions are channeled and propagated via social media into a network of people following/consuming it. The research and the proposed model can, however, be extended to other social media and be analyzed for a bigger social network.

The study is performed in three stages. In stage one, subjectivity versus objectivity in news is investigated. Ideally, media outlets should report neutral and objective news to its consumers. However, investigations reveal that a large fraction of the them report subjective news to its consumers, thereby injecting opiniated information into the network. This subjective news can be referred to as biased opinion.

In stage two, the short term impact such subjective news reporting has

* Corresponding author.
E-mail addresses: swati1178@gmail.com (S. Aggarwal), sinha.tushar94@gmail.com (T. Sinha), yashkukreti8117@gmail.com (Y. Kukreti), siddarthshikhar@gmail.com (S. Shikhar).

[1] Equal contribution.

https://doi.org/10.1016/j.array.2020.100025
Received 29 September 2019; Received in revised form 3 April 2020; Accepted 8 April 2020
Available online 18 April 2020

on its consumers is studied. The proposed work aims to analyze how media outlets subtly condition news consumers in both positive and negative overtones i.e. the news consumers may get conditioned to be in line with the media outlet's view on a subject, or may get conditioned to have a completely opposite view with respect to the media outlet's view. It is evident from the results obtained that large number of news consumers develop biases towards a particular issue, as guided and conditioned by media outlets.

Finally, having justified the need for a platform to receive safe and healthy objective news, a system is proposed to detect alarming media biases by measuring their short term impact on news consumers. A unique 'bias short term impact score' calculation method is proposed which leverages social media metrics as extracted from media outlets and their followers coupled with polarity scores as obtained from stage one. To detect alarming biases, a subsequent comparison of this score with a neutral baseline is performed and this forms the third stage of this research.

The paper presents a comprehensive description of these three tasks in sequential order.

2. Literature survey

The task of sentiment analysis is central to all the three stages of this work. It forms the basis to classify and quantify the polarity of tweets in the dataset. Therefore, this task needs to be carried out in the most efficient manner.

Traditional approaches to sentiment analysis rely either on lexicon-based approaches or implementations through machine learning algorithms.

Liu [2] along with Pang and Lee [3] provided a comprehensive review of the methods used in sentiment analysis. They discussed the usage of unigram features like term frequency, POS tagging, syntax analysis, negation handling and domain considerations while using lexicon-based models. They further discussed supervised approaches by feeding these features to Naïve Bayesian or SVM (Support Vector Machines) classifiers. For unsupervised methods, PMI calculations, which measure the closeness of two words, were also examined to determine the sentiment score.

Following this, Hutto and Gilbert [4] constructed a gold-standard sentiment lexicon – VADER (Valence Aware Dictionary and sEntiment Reasoner) which was specifically attuned to microblog-like contexts. As part of their work, they created a rule-based model to consider the grammatical and syntactical conventions used by humans to emphasize sentiment intensity which could be implemented irrespective of the underlying lexicon/algorithmic implementation. VADER was trained on a dataset consisting of tweets and outperformed human raters. It generalized more favorably than any benchmark including other lexicons or machine learning techniques.

VADER, in its comparison to lexicons which provide a binary classification of words like LIWC and GI, showed better performance in the social media domain due to better capturing of social media specific lexical features and a tendency to account for the sentiment intensity of words rather than just their binary nature. It even outperformed lexicons consisting of a sentiment intensity value mapping like ANEW and SentiWordNet, due to lexical feature capturing for the social media domain, and being less noisy due to its gold standard i.e. human curation.

Its performance was comparable to machine learning approaches involving SVM's and Naïve Bayes classifiers, however, it had an added advantage of not requiring a training dataset, not being overly reliant on the training dataset validity for efficient feature extraction and not requiring computationally expensive processes, something, which is a huge positive while considering the real-time nature of social media.

Due to all of the above reasons accompanied with its ease of usage, VADER was identified as an apt fit for the use case of this work. Therefore, this proposed work uses the VADER tool in all further tasks for calculating the sentiment scores for tweets in the dataset.

2.1. Investigating subjectivity vs objectivity for news

In the last decade, there has been substantial research analyzing the role of news coverage by media outlets.

Garon [5] studied the functional subjectivity in media coverage, specifically focusing on The Gulf War case. Prime time news reports were obtained from news channels like CNN, TV5, CBC and CBV for a period of 28 days. This was followed by systematic content analysis on each news story with subjectivity assessment done by qualitative indices like the degree of portrayal of censorship and presence/absence of actual sources to back up stories. Following probability statistics standards which identified subjectivity from closely related Chi-square values, the relationship between subjectivity in television news reports and the media's influence strategies were determined.

Similarly, White [6] investigated news subjectivity by referring to a Sydney Morning Herald report on a violent attack after the Gulf War. The rhetoric potential of modern English language news items was studied by observing that the presentation of news which seemed 'objective' according to media's ideals of 'neutrality', 'balance' and 'reliability', still conveyed social evaluations and interpretations based on the author's personal preferences. The media's claim for objectivity of text was simultaneously addressed by analyzing different sections of the news article and discussing the implications of an orbital model used for reporting 'hard news'.

Lex and Juffinger and Granitzer [7] assessed objectivity in online news media by analyzing articles from two British newspapers, 'The Telegraph' and 'The Guardian'. Topic independent features were used with standard bag-of-word models to classify articles as objective or subjective. It was observed that topic independent features resulted in a gain in accuracy for cross-domain experiments, however, for single-domain experiments, the standard bag-of-words model without stopword removal and stemming provided optimal performance.

This proposed work uses a simple, computationally inexpensive method to identify subjectivity in news tweets. It is different from the aforementioned researches as it uses an unsupervised computing approach. The computational efficiency of the proposed work is attributed to the simplicity of unsupervised methods over the complex supervised approaches (which include annotating large datasets and computational expense of supervised algorithms). Also, since the analysis is to be performed on a single domain, topic independent features, like the ones implemented by Lex [7]; are not required. The subjectivity of news is determined by checking whether the sentiment score (obtained from VADER [4]) lies within pre-established limits.

2.2. Impact of reporting on news consumers

Various researches have studied the impact of subjective news reporting on its consumers.

O'Connell [8]; for instance, studied the reasons for fear among the citizens of Ireland of a 'law and order crisis', as visible from surveys, despite low crime rates and related the cause of this misconception to the distorted image of crime in media. Articles on crime were collected from four leading newspapers for a period of two months. Parameters like date, newspaper, page number, headline, number of words, area of article and the type of offence reported in the article were recorded. These parameters were used to account for the ways in which the media induces bias.

DellaVigna and Kaplan [9] analyzed the entry of Fox News in cable markets and its impact on voting. Using a dataset consisting of voting data and Fox News viewership metrics in local cable channels, its effect on the vote share in Presidential elections of 1996 and 2000 was calculated. It was observed that Republicans gained 0.4 to 0.7% points in the towns that broadcasted Fox News; concluding that Fox News convinced 3 to 28% of its viewers to vote for Republicans. The vote share change was attributed to the media slant induced by Fox News by mentioning results of 2 previous experiments which showed that 'evaluators of information

do not take sufficiently into account the (known) incentives of the advisors and are thus persuaded by their advice' and 'small investors follow the recommendations of affiliated analysts, despite the conflict of interest of the analysts.' The proposed work in this paper banks on a similar explanation to show conditioning among the public when they consume subjective news tweet made by a prominent media outlet.

Gerber and Karlan and Bergan [10] explored the effects on political views and behavior of media bias by conducting a natural field experiment. Individuals were randomly assigned to either receive a free subscription to the Washington Post, Washington Times, or to a control group which did not receive any newspaper. Door-to-door polling models were used followed by probabilistic analysis. A public opinion survey was conducted after the 2005 Virginia gubernatorial election to find that irrespective of the media slant, people who received newspapers had a greater tendency to vote Democrat. Also, a long-term increase in the voter turnout was observed in the 2006 election, possibly, due to re-subscription of these newspapers. Finally, it was inferred that not only media slant, but even media exposure, can cause an evocation in public opinions.

Chiang and Knight [11] also investigated the influence of media on voting in the context of newspaper endorsements by developing a statistical econometric model in which voters chose candidates by resolving their uncertainty over the quality of candidates by relying on endorsements from newspapers. The degree of influence was found to be dependent on the credibility of the endorsement. This observation led to an interesting side-effect of the influence phenomenon where voters were found to be sophisticated enough to filter out evident biases in media, while still retaining some orientation with endorsements published in credible newspapers.

This proposed work introduces an approach to determine if news consumers have been conditioned, by comparing the current sentiment value of the consumer with its previous average sentiment value. This difference is then compared against a threshold to determine the extent and nature of conditioning. If considerable change is noted, the news consumer is said to have been conditioned, either positively or negatively.

2.3. System to detect alarming media bias and measure its impact

Many theories have been proposed to measure and quantify inclinations and biases of media outlets.

Groseclose and Milyo [12] provided an objective measure of the slant of news and a comparative measure against other political actors. Americans for Democratic Action (ADA) scores, which categorize user sentiment in the range 0 (strongly conservative) to 100 (strongly liberal) were estimated for all major media outlets in the US by counting the number of times a media outlet cited various think tanks. These citation patterns were compared with the number of times the Congress cited the same think tanks in their speeches on the floor of the House and Senate. The estimation was also dependent on the pre-existing adjusted ADA scores of the parliamentarians making these speeches. Finally, ADA scores were calculated for each media outlet on 2 granularity levels: sentence level, and citation level.

Lin and Bagrow and Lazer [13] proposed measures to quantify the extent and dynamics of blog and mainstream media bias by focusing on stories about the 111th US Congress. A networked data model was created using different node sets for mainstream media, blog media and legislators. The observed coverage of the Members of Congress was then compared against a null model of unbiased coverage and biases with respect to political parties, popular front runners, regions of the country, gender and ideology were investigated. Media slant was also observed in the content of the coverage by examining links used in the text along with sentiment analysis on the textual content.

An et al. [14] proposed a model to map the news media sources along a one-dimensional political spectrum using Twitter co-subscriptions relationships between them. The relative position of each media source was determined on a one-dimensional space by considering its distances to other media sources. The distance was calculated by observing the common subscriber patterns. The final co-ordinate was calculated as an ADA score on the basis of its distance to some landmark media outlets, whose coordinates were already known with a high degree of confidence as calculated by Groseclose and Milyo [12]; which was then plotted on the one-dimensional space.

Dunham [15] examined the political bias in six large daily newspapers. More than 25,000 references made to 12 think tanks over 18 years and 10 ideological frames (such as liberal, conservative or libertarian) were examined. The rate of attachment of these ideological frames to think thanks by the means of news articles in the dataset was then investigated to calculate the biasness of the media source. The media examined were found to be more likely to associate ideological labels with organizations or think tanks with a conservative orientation, than with those having a liberal orientation, thereby, suggesting a liberal bias.

This proposed work introduces a novel system to detect media biases and measure their short term impact on news consumers. The system forms an integral part of the platform for news consumers to receive safe and healthy objective news. Various parameters like influence, reach and persistence of a tweet are used to calculate the Bias Short term Impact Score (BSIS) (using a formula further explained in Section 5.3.3) which measures the 'bias-ness' of tweet. The parameters are calculated by quantifying standard Twitter actions like retweets, likes, replies and mentions.

3. Subjectivity vs Objectivity in news

It is the responsibility of media outlets to report day to day events in an unbiased and neutral manner. However, bias in selecting what to report and concomitantly choosing a slant on a particular report plagues the current system. This study specifically focuses on Indian media outlets, relying on Twitter datasets to represent biased reporting and proposes an approach to detect it.

As mentioned earlier in Section 1, the investigation can be extended to media other than Twitter as well. The domain specificity to Twitter has been chosen for simplicity.

Journalists with a large following tend to be quite influential, with their opinions getting propagated to the masses. They can strongly affect public opinions with their influence having a strong correlation with the subjectivity of their tweets. This may lead to a precarious situation where mass opinion is dependent on the whims and personal orientations of few journalists and can therefore alter the political state of the country. It is therefore imperative to identify such journalists to restore the concept of free will in society.

The study first validates the assumption that media outlets provide the news consumers with opinionated news. Instead of providing fair, neutral - objective news, it is shown how various media outlets and journalists provide subjective news. To do so, the tweets by popular media outlets and journalists are analyzed and checked for news subjectivity versus objectivity. It is found that a large fraction of the tweets are subjective instead of the ideal objective nature.

3.1. Data description

Twitter is a popular social networking and microblogging service that allows users to post real time messages, called tweets. Tweets are short messages, restricted to 140 characters in length at the time of the study).

With the rise of social media in India, there has been an increase in the number of people using Twitter. Almost all media outlets and journalists have verified Twitter handles, which they use periodically to tweet information about various day to day happenings and events in the country.

One such event was the Demonetization move by the Government of India (GoI). On November 8, 2016, the GoI announced the Demonetization of Rs. 500 and Rs. 1000 currency notes to curb the menace of black

money. This provoked wide-ranging reactions from people, with some appreciating the move for its noble objectives, and some criticizing it citing the massive inconvenience that followed the announcement.

The Twitter API [16] is used to fetch real time tweets from the verified handles of 10 media outlets and 10 journalists. A total of 4475 tweets are collected for the subject of "Demonetization" and subsequent analysis is done. (Tweets were extracted for the period Nov 8, 2016–Feb 16, 2017).

Note: As part of this research, data was extracted for two other major events in the Indian context. These were:

1. Odd-Even Rule in Delhi 2017 (Private vehicles could be on the road only on certain days, depending on their license plate number, from January 1, 2016. The scheme aimed to reduce pollution and smog in Delhi)
2. Uttar Pradesh Assembly Elections-2017 (Election to the 17th Uttar Pradesh Legislative Assembly)

Upon preliminary investigation following data pre-processing (as mentioned in Section 3.1), these datasets were found to be unfit for evaluation. This was primarily because:

1. Both of these events were not as significant on a national scale as compared to demonetization. Demonetization was a nation-wide event which invoked wide-ranging reactions from media outlets as well as citizens across the country.
2. Due to the regional nature of the two events, it was found that the tweets were noisy due to presence of larger proportion of regional media outlets as compared to the set of media outlets studied through the proposed work. Post cleaning, such tweets did not have enough context for further research.
3. The period over which these events invoked reactions led to a much smaller dataset as compared to the studied dataset. Both the events studied did not have long-ranging life in the media.

3.1.1. Pre-processing

The raw tweets fetched in Section 3.1 contain noise in form of language discrepancies, improper formatting, improper punctuation, use of medium (here Twitter) specific symbols and emoticons etc. This warrants some pre-processing in order to clean the tweets and make them suitable for further computation.

The steps carried out for pre-processing are enlisted as:

- All URLs are replaced with a tag "URL"
- All targets (For e.g. "@narendramodi") are replaced by the username and dropping the '@' at the start (e.g. gets changed to "narendramodi").
- All punctuation marks occurring in the text at the start or end of a word are separated from the word to enable easier tokenization and therefore better performance (For e.g. said' gets changed to said ').

3.2. Polarity score

Sentiment scores for the tweets are determined by analyzing the tweets collected using VADER. NLTK- VADER is a popular tool for sentiment analysis of social media texts. It comprises of an underlying implementation of a rule-based model (Hutto and Gilbert, 2014). Written in Python, the library is one of the most commonly used tools for determining the polarity or sentiment score for given text in natural language. The sentiment scores obtained by VADER are used for determining the polarity of text.

VADER provides a polarity score (*PS*) ranging from −1 (strongly negative) to +1 (strongly positive). This range is split into 3 regions, namely positive, negative and neutral. *PS* values ranging from −1 to −0.33 are classified as negative *PS* values while *PS* values ranging from 0.33 to 1 are classified as positive *PS* values. *PS* values lying between −0.33 and 0.33 are referred to as neutral *PS* values. The visualization is shown in the number line plotted below in Fig. 1.

The division of the polarity score space into three distinct sections ensures that tweets which are significantly polar are the only ones classified as positively or negatively biased. It is possible that genuine tweets on Demonetization, that are fairly neutral, have a non-zero PS value. A PS between −0.33 & 0.33 is only slightly polar, and not necessarily biased. This reduces false positives while identifying biased tweets.

The pseudocode for determining polarity of text is given as below in Fig. 2.

3.3. Observations

Contrary to the ideal objective news, a large fraction of tweets by media outlets/journalists are observed to be polar or subjective in nature. Table 1 below shows the observations on the dataset. Fig. 3 then shows these observations in graphical format.

Thus, it is concluded that instead of providing neutral/objective news, the tweets by media outlets are largely subjective in nature.

4. Conditioning

As shown by means of the aforementioned investigation, media outlets and journalists present news consumers with polarized opinions. This guidance and conditioning through subjective news articles can be asserted as the cause for news consumers to become biased towards a particular issue. This section of the study proves this assertion and analyzes the conditioning of opinions by media outlets.

4.1. Terminology

Before describing the method to analyze conditioning of opinions, a few terms are introduced below.

4.2. Data description

The polar tweets, as extracted from the overall tweet dataset used in Section 3, forms the basis for analyzing conditioning by media outlets. Along with the data as mentioned in Section 3.1, for each polar tweet, a list of user-ids of the users who 'react' to the given tweet is also extracted. The behavior of 'reactors' to polar tweets as obtained from Section 3 is observed to study the conditioning of opinions by the media outlets.

4.3. Methodology

The conditioning of news consumers is analyzed by observing their activity prior to and post the event (*E*). Of all the users who consume/view the polar news/tweet, the study relies on the 'reactors' of the given tweet as mentioned in Section 4.1 and their behavior is investigated as explained before.

Each user has prior opinions on the issues that are happening around him.. The prior polarity (*p*), as defined in Table 2, of a person is calculated by the exponential moving average (*EMA*) of the polarity score (*PS*) of its previous tweets. **Exponential Moving Average (*EMA*)** for the *PS* series is calculated recursively as:

$$p(t) = a^{*}PS(t) + (1-a)^{*}p(t-1) \quad (1)$$

The value of a is chosen as 0.3. This is done to provide fair weightage to both the last tweet and the series of previous tweets.

However, a polar tweet/news report by a media outlet can affect the user's opinion on the issue. The polarity in the opinion of an user might show a massive swing (reversing one's opinion) or a minor one (a small strengthening of prior opinion). Various other external events/happenings can be a reason for this opinion change, but the time limit of t to

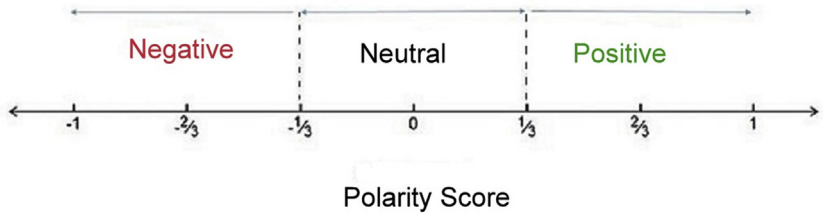


Fig. 1. Range of negative, neutral and positive polarity sentiment scores.

- Store tweet text in string str
- Run NLTK-Vader on str and store sentiment score as score
- if score<-0.33
- Categorize tweet as **negatively subjective**
- Else if score>=-0.33 and score<=0.33
- Categorize tweet as **neutral**
- Else
- Categorize tweet as **positively subjective**

Fig. 2. Pseudocode for determining polarity of text.

Table 1
Polarity distribution of tweets.

Number of tweets	4475
Number of polar positive tweets	873
Number of polar negative tweets	931
Percentage Of Polar Positive tweets	19.50%
Percentage Of Polar Negative tweets	20.80%

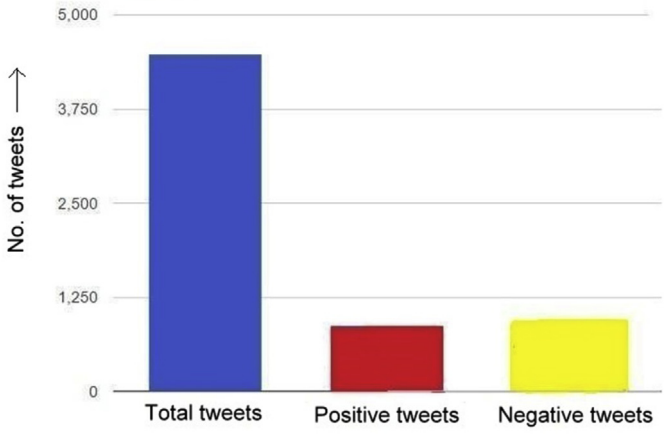


Fig. 3. Graphical representation of the polarity of tweets in the dataset.

t+24 h is taken to ensure that the swing in opinion can be attributed largely to the polar news consumed by the user.

The post-event polarity (p'), as defined in Table 2, is obtained by taking the arithmetic mean of the polarity scores of its tweets post the event for a period of 24 h. It is assumed that the impact of conditioning is diluted and can not be attributed with sufficient confidence to twitter data after 24 h.

An event E is said to have conditioned a consumer C if the value of p and p' are significantly different. The two opinions p and p' are assumed to be significantly different if the difference of p and p' is 20% more (1.2 times) than the standard deviation in the PSs of C prior to event E. The factor of 20% is taken to ensure that the change in polarity is significant, and not just an impulse in the reactor's PS. Only a change large enough with respect to the standard deviation is considered to be associated with

Table 2
Terminology used in the paper.

Term	Definition
Polarity Score (*PS*) of a tweet	As defined in Section 3.2, it denotes the sentiment value of a tweet. The score ranges from a scale of −1 (strongly negative) to +1 (strongly positive).
Event (*E*)	A subjective tweet by a media outlet, i.e. a tweet having polarity score (*PS*) values between −1 and −0.33 depicting a polar negative tweet or between 0.33 and 1 depicting a polar positive tweet, is referred to as an event. The time at which the event occurs is denoted by *t*.
Consumer (*C*)	A tweet by a media outlet or a journalist is fed to all followers of its Twitter handle. These followers are referred to as the consumers or target audience of the tweet.
Prior polarity (*p*)	Each news consumer has a prior polarity indicating its general prior opinion on a particular person/topic.
Post-event polarity (*p*')	The post-event polarity of a news consumer indicates its opinion on the particular person/topic after an 'event' occurred. .
Reaction	A Reaction to a tweet is defined as engagement in any of the following activities to the tweet: Like, Reply, Retweet. The detailed definition of these reactions and their impact is discussed in later section.
Reactor	A user who has 'Reaction' to the tweet under consideration.
Active Reactor (*AR*)	An 'active reactor' is defined as a 'reactor' who tweets within 24 h of the polar tweet to which it reacted.

the journalist's tweet and, therefore, fit to be labelled as an act of 'conditioning'.

Note that an influence can be both positive and negative. An event E which causes a consumer C to be induced with a polarity change in line with that of the tweet in E is said to have had a positive influence on C. Similarly, an event E which causes the consumers C to be induced with a polarity change opposite to that of the tweet in E is said to have had a negative influence on C.

The pseudocode for this process is shown in Fig. 4 below.

The entire process can be summed up and represented using a flow chart as shown in Fig. 5. It shows the various activities that are carried out to examine the conditioning of news consumers by the media outlets/journalists.

4.4. Observations

The tweets of 'reactors' are analyzed to study their behavior before and after the media outlets' polar tweet event. In the proposed study, tweets made by the Top 25 'Retweet-ers' and 'Favourit-ers' of a specific polar tweet by a media outlet/journalist are analyzed. These Top 25 reactors are extracted with the help of the Twitter API [16] and web scraping techniques. Table 3 below shows the statistics regarding these tweets:

The tweets of these AR's, as defined in Table 2, are analyzed to investigate if and how they are conditioned. This act of conditioning can be positive or negative as mentioned previously in Section 4.3.

A polar tweet which 'conditions' the AR such that the direction of swing in polarity for its tweets in the next 24 h is in line with the polarity of the actual tweet, it is said to have positively conditioned the

```
-Fetch all the tweets of a consumer for period to be analysed
-Initialize PS,p,stddev arrays with size as no. of tweets and
 all elements initialized to 0
-For each tweet
-       Extract sentiment score of tweet
-       Store sentiment score in corres. PS array slot

-For each tweet
-       Calculate EMA for past tweets and store values
        in corres. p array slot by using formula
        p[i]=0.3*PS[i]+0.7*p[i-1]

-For each tweet
-       Calculate standard deviation values from first tweet
        upto current tweet and store in corres. stddev array slot
        by formula stddev[i]=sqrt((PS[i]-p[i])*(PS[i]-p[i])/(i))

-Extract tweets (from already existing tweet array) made in time
 t to t+24 Hours and assign starting index of this sub-array to
 start_ind and ending index of sub-array to end_ind

-Find mean of PS values for sub-array ranging from PS[start_ind]
 to PS[end_ind] and assign mean to mean_PS

-if abs(mean_PS-p[start_ind-1])>=1.2*stddev[start_ind-1]
-       Infer that consumer has been conditioned
-else
-       Infer that consumer has not been conditioned
```

Fig. 4. Pseudocode for determining whether a consumer has been conditioned.

AR (*PCAR*) and if the direction of swing in polarity for its tweets is opposite to the polarity of the actual tweet, it is said to have negatively conditioned the *AR* (*NCAR*).

The results obtained in this task are shown in Table 4.

13.39% of the reactors analyzed are classified as active reactors. Among the active reactors, 41.23% reactors are found to be conditioned. Incidentally, in this case study the number of positively and negatively conditioned reactors turn out to be almost equal, occurring in a 52–48 ratio.

4.4.1. Sample PCAR observation

An example of how an active reactor is classified as a positively conditioned active reactor is shown below:

Example 1: Journalist's Tweet after pre-processing: 'Day 9: Demonetization Death Toll Rises To 55 by DilliDurAst URL'.

Step 1: Sentiment score of journalist's tweet (as obtained from VADER (Hutto and Gilbert, 2014)): *0.5994*
Step 2: Reactor's tweet after pre-processing: *'The savage impact Modi's haphazard demonetization is having on the economy ….'*
Step 3: Sentiment score of reactor's tweet (as obtained from VADER): *0.4588*
Step 4: Previous average sentiment score of the reactor: *0.296*
Step 5: Difference between average sentiment and current sentiment: *0.1628*
Step 6: Previous Standard Deviation of sentiment for reactor: *0.05*

Since the difference is more than 1.2 times the standard deviation, and the change in polarity is in line with the polarity of the actual tweet, it is inferred that the reactor is positively conditioned.

4.4.2. Sample NCAR observation

An example of how an active reactor is classified as a negatively conditioned active reactor is shown as:

Example 2: Journalist's Tweet after pre-processing: *'Our coverage and big focus on demonetization panic on streets continues: coming to you from*

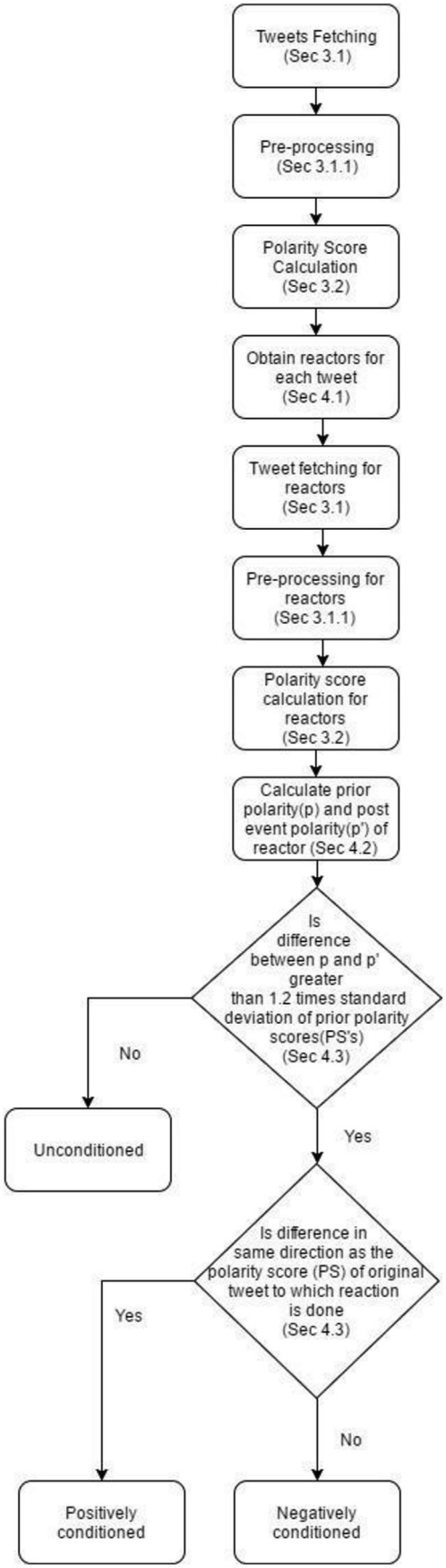


Fig. 5. Flow diagram for determining conditioning.

Table 3
Reactor sample Data Size.

Number of tweets analyzed from 'reactors'	116212
Number of 'reactors' analyzed	8350

Table 4
Data summary of reactors and their tweets.

Number of 'active reactors'	1118
Number of conditioned 'active reactors'	461
Number of positively conditioned 'active reactors'	240
Number of negatively conditioned 'active reactors'	221
Number of Positively conditioned tweets from 'active reactors'	370
Number of Negatively conditioned tweets from 'active reactors'	376

icici at sector 18 Noida:lines don't end'.

Step 1: Sentiment score of journalist's tweet (as obtained from VADER (Hutto and Gilbert, 2014)): *0.5106*
Step 2: Reactor's tweet after pre-processing: *'That is true. He's also helped by poor opposition to #Demonetization . Regardless, he has a winner.'*
Step 3: Sentiment score of reactor's tweet (as obtained from VADER): *0.5423*
Step 4: Previous average sentiment score of the reactor: *0.8074*
Step 5: Difference between average sentiment and current sentiment: *1.3497*
Step 6: Previous Standard Deviation of sentiment for reactor: *0.122*

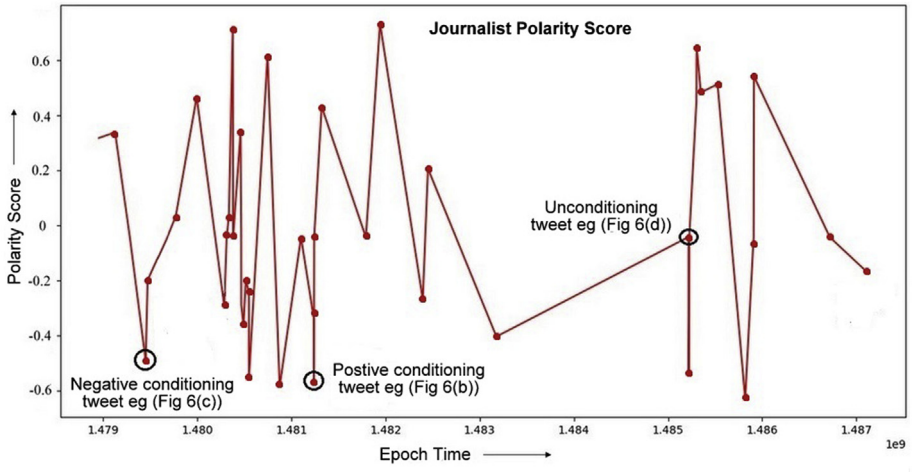
Since the difference is more than 1.2 times the standard deviation, and this change in polarity is opposite to the polarity of the actual tweet, it is inferred that the reactor is negatively conditioned.

4.4.3. Graphical analysis/observations

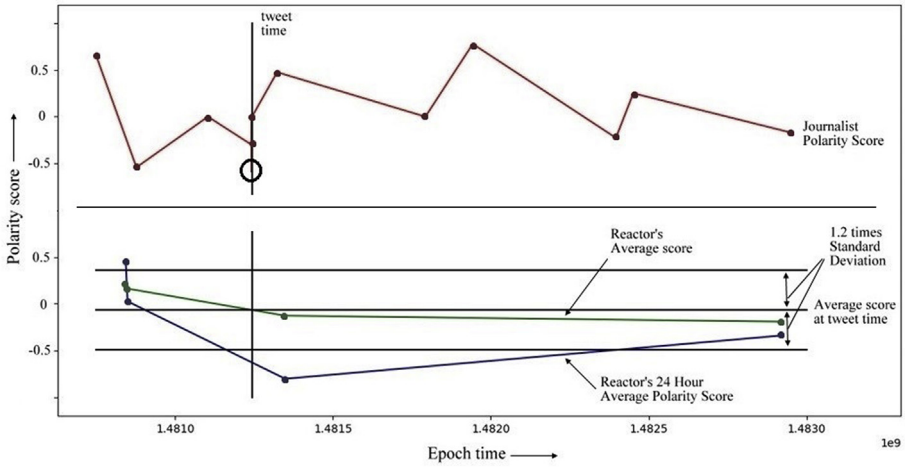
The analysis of some more test cases is shown in the form of the following graphs in Fig. 6.

Fig. 6(a) shows the polarity score plot (red line) of a journalist with time. The three points encircled in this plot correspond to the example tweets for different types of conditioning. Fig. 6 (b), (c) and (d) zoom into the polarity score plot at these points, and show the details of conditioning due to the corresponding tweets. These include the 24-h average of the polarity score (blue line), the exponential moving average (*EMA*) of polarity score plot (green line), of a 'reactor' who is conditioned/not conditioned by the journalist's tweet and the journalist's polarity score plot (red line). The 24-h average is calculated as the average of the polarity score of all tweets by the reactor in a 24-h period post a tweet. The EMA is calculated using (1). The vertical line marks the tweet time in the graph.

If the difference between the 24-h average of the polarity score plot (blue line) and the *EMA* of polarity score plot (green line) at tweet time is greater than threshold, we say that the reactor is conditioned. In

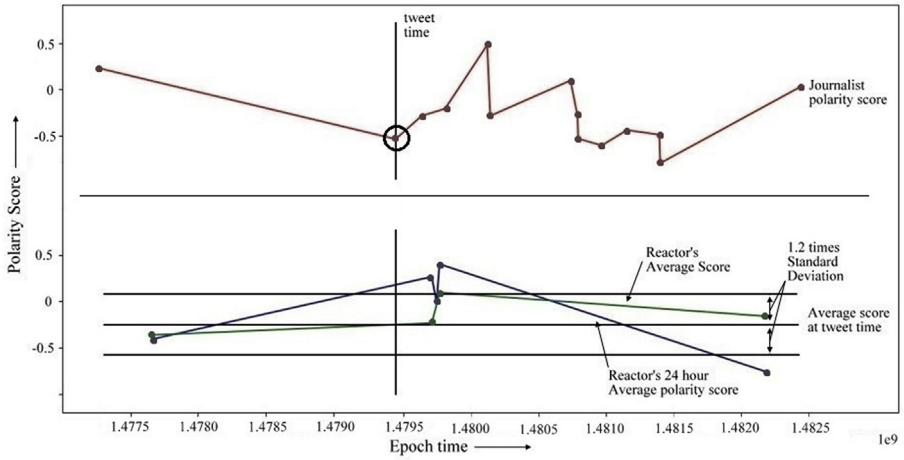


(a) Journalist polarity score plot

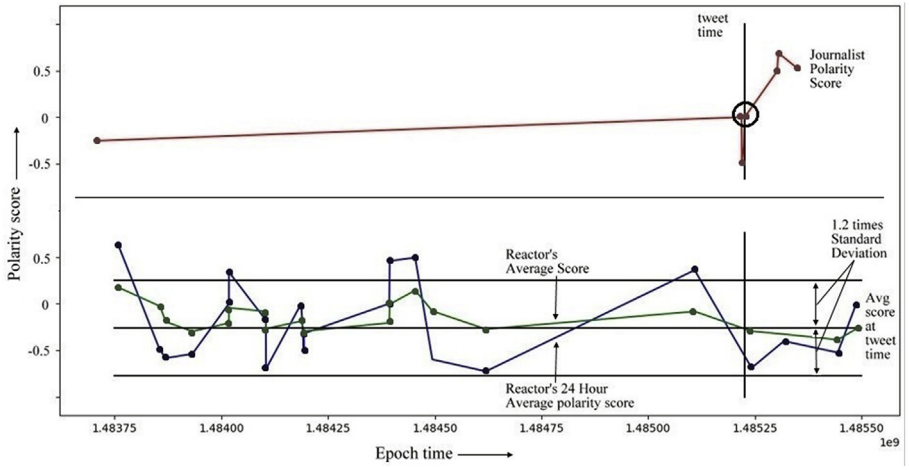


(b) PCAR illustration

Fig. 6. (a) Journalist polarity score plot. **6(b)** PCAR illustration. **6(c)** NCAR illustration. **6(d)** Unconditioned *AR* illustration.



(c) NCAR illustration



(d) Unconditioned *AR* illustration

Fig. 6. (*continued*).

Fig. 6(b), the graph drops by a value greater than the prescribed limit after a negative polarity tweet by the journalist, hence the reactor is said to be positively conditioned (PCAR). In Fig. 6(c), the reactor's graph again drops below the calculated threshold, however, this fluctuation occurs following a positive tweet by the journalist. Hence, this reactor is said to be negatively conditioned (NCAR). Finally, in Fig. 6(d), the reactor's graph remains within limits on both sides of the average polarity score plot. Therefore, this reactor is not conditioned by the journalist's tweet.

5. Bias short term impact assessment

The study conducted so far shows that a large fraction of tweets by media outlets are polar instead of the ideally expected neutral nature which, on further evaluation, is found to have the power to affect news consumers. These findings justify the need for a system which can detect alarming biases arising from news articles in order to provide the base for future corrective action.

Next, a model is proposed to detect and measure these biases. The bias short term impact assessment tool discussed in this section is an integral component needed to build a platform for safe and objective news. It implements a unique bias score calculation mechanism considering a diverse set of factors followed by comparison with a neutral baseline.

5.1. Data description

The overall tweet dataset used previously in Section 3.1 is extended by using the tweet permalinks and user profile pages to extract other useful information about each tweet. This information includes the number of followers of the user, the number of times the media outlet is mentioned on Twitter, the number of likes, replies, retweets of the tweet, the activity time of tweet, user-details of the users replying to the tweet and text-content of the tweet. These fields are used for determining the short term impact of bias as described below.

5.2. Polarity and subject identification

Along with the polarity score, the subject or feature for which the sentiment is attributed is also determined. This task is carried out by using feature identification strategies as suggested by Pang and Lee [4].

For experimentation, some major political subjects were identified for the Indian context. This included the Bharatiya Janta Party (BJP) (Current Ruling Party at the Centre of the Indian Government), the Indian National Congress (INC) (Leading Opposition Party at the Centre) and the Aam Aadmi Party (AAP) (Current Ruling Party in the Delhi Legislative Assembly) and their prominent leaders. The subjects/features were hand-annotated and grouped into discrete sets called target sets.

Now for each tweet, along with the sentiment score, the target at which the sentiment is directed is also identified. The target is one of the above-mentioned sets. Linguistic features like POS (Part of Speech) tags

and NER (Name Entity Recognition) tags are used for subject identification in the tweet text.

This is followed by rudimentary checks to identify the subject for each tweet. It is observed that most tweets belong to just one target set. This implies that complex methods like dependency graph representations to identify the subject and target sets can be safely avoided. Presence of unique target sets allow simpler methods to be implemented.

The polarity score of the tweet is assumed to be an opinion on the subject identified.

5.2.1. Illustration

Example 3: A tweet with the text *"PM asks BJP MPs & MLAs to give details of bank accounts from Nov 8 (date of demonetization) to Dec* 31st . *Huge Bold move. Party* funds *next?"* has a VADER (Hutto and Gilbert, 2014) polarity score of 0.765 indicating high positive sentiment.

'BJP' is identified as the subject by the proposed model and the sentiment score obtained from VADER is attributed towards the target set 'BJP'.

Example 4: Similarly, a tweet with the text *"The man who should speak on Demonetization Dr Manmohan Singh hasn't said a word. Guess there is too much noise elsewhere these days!"* has a VADER polarity score of −0.21 indicating a negative sentiment.

'Manmohan Singh' is identified as the subject by the proposed model and the sentiment score obtained from VADER is attributed towards the target set ('Indian National Congress' in this case) to which 'Manmohan Singh', an Indian politician, belongs.

5.3. Bias and its short term impact

Any news with a polarity score outside the neutral limit (−0.33 to +0.33) is assumed to be subjective or biased as mentioned in Section 3.2. However, not every subjective news will have the same impact. The impact of subjective news reporting, depends on various other parameters like the number of people who view the news, number of people who like/favor it, and the time for which the news is active in the user newsfeed. Depending on these parameters, the Bias Short term Impact Score (*BSIS*) is determined for each news report (using a formula explained further in Section 5.3.3).

A *BSIS* timeline is then constructed for each media outlet to analyze the biased reporting and its associated short term impact factor. A bias of high impact score is recorded as an alarming bias, as explained later in Section 5.3.3.

5.3.1. Factors affecting BSIS

The proposed work models the solution specifically for the Twitter domain. However, the factors listed below have generalized analogues in other media as well, and hence, the model can be extended for media other than Twitter easily. The various factors affecting the Bias Short term Impact Score (*BSIS*) are listed below:

1. **Direct Followers (*F*)**: The number of users who follow the media outlet represents the number of news consumers. It is an approximation for the audience that will view the tweet. It is extracted with help of the Twitter API [16].
2. **Mentions**: The number of times a media outlet is mentioned in tweets by other people in the 'Twittersphere' (the entire microblogging Twitter world is referred to as a Twittersphere) is indicative of how actively others look up to it. It abstractly measures the popularity of the media outlet. The total number of mentions of a media outlet in the period of observation are determined and used as a measure of their popularity. This is called the number of 'overall mentions'.

In contrast to overall mentions, the term 'immediate mentions' is used to capture the number of mentions of a media outlet in the time span 24 h prior to and 24 h post the tweet. This interval measures how actively the media outlet was mentioned prior to the tweet (In the time span between $t-24$ to t, where t is the time of tweet in hours) and post the tweet (In the time span between t to $t+24$, where t is the time of tweet in hours). Whilst the first term captures the popularity of the media outlet, thereby determining the audience of the tweet, the latter determines the short term impact of the tweet by measuring the mentions (which in turn captures the audience reaction) post the tweet.

Both varieties of 'mentions' are extracted by using web scraping techniques.

3. **Tweet Reactions**: The news consumers may choose to react to a particular tweet using standard Twitter reactions. They are referred to as 'reactors' in Section 4.1. The impact factor of their actions can be defined in detail as –
   a) **Like:** A user's liking/favoring a tweet indicates their appreciation/approval for the tweet sentiment. The Top 25 'Favourit-ers' are extracted for each tweet by using standard web scraping techniques.
   b) **Retweet:** A retweet/share indicates even higher level (as compared to a 'Like') of appreciation/approval as the user chooses to share the tweet from its own handle indicating a strong alignment with the media outlets' view. The Top 25 'Retweet-ers' are extracted for each tweet by using the Twitter API [16].
   c) **Replies:** User replies may either agree or disagree with the media outlet. The agreement/disagreement is determined by sentiment score comparison of the tweet and reply. The subject identification methodology, as suggested in Section 5.2, is also used as a parameter for this decision. An agreement is defined as the case when polarity of the tweet and reply are the same (for the same target subject). Similarly, a disagreement is classified as the case when polar nature of the tweet and reply conflict with each other.

The replies to a tweet along with the Twitter handle of the user making the reply are extracted for each tweet by using web scraping techniques.

4. **Tweet Activity Time**: The time interval for which a tweet is active is a measure of its short term impact. If a tweet by a media outlet remains active for a long duration, it implies that the tweet elicited a good response from the target audience and hence the impact of the news bias was sizeable. The tweet activity time is measured by finding the time difference between the time of tweet and the last reply on the tweet. The Twitter API [16] is used to obtain the timestamp of a tweet.
5. **Indirect/Induced Followers**: Along with the direct followers, the audience of a tweet also includes a number of people who indirectly view the tweet as a result of someone whom they follow. For example, if a user *A* retweets a user *B*, then the tweet by user *B* automatically gain the followers of *A* as its audience. Similarly, other tweet reactions (like and replies) induce indirect followers as they cause the tweet/post to appear on the user's newsfeed. It is calculated by summing the number of followers of all user-id's who react to a tweet.

5.3.2. Terminology

Before describing the formulae for calculating *BSIS*, some terms are introduced below.

Note: Due to lack of labelled datasets and highly subjective nature of biases, supervised approaches cannot be applied for this use case. These formulations use a statistical approach where the value of constants (parameter weights) and operators used in the formulations are based on theoretical investigations. The formulae may be improved in the future by taking various other factors which might not have been included presently into account.

It should be noted that this research focuses more about the terms and the meaning they represent, leaving the improvement of mathematical formulations as a future to-do task.

1. **Polarity Score (*PS*) of a tweet**: As defined previously in Section 3.2, it denotes the sentiment value of a tweet. The score ranges from a scale of −1 (strongly negative) to +1 (strongly positive).
2. **Influence Factor (*IF*) of a media outlet:** A measure of how influential a particular media outlet is. It is a function of the direct followers and number of mentions of the media outlet in the Twittersphere. Let *F* be the number of followers and *M* be the number of overall mentions of a media outlet. These two parameters together determine the popularity of the media outlet. The value of *F* directly measures the audience viewing a tweet by the media outlet, and *M* measures how actively people engage the media outlet in their discussions. *IF* is given by:

$$IF = M*F \tag{2}$$

Note that the operator * aggregates the values of *M* and *F* together and best captures the influence as *IF*.

3. **Reach (*R*) of a tweet:** Measures the coverage or reach of a tweet as a function of the following parameters:
   - (*rt*): Number of retweets (Refer 3(b), Section 5.3.1)
   - (*fav*): Number of favorites (Refer 3(a), Section 5.3.1)
   - (*am*): Number of users who replied in agreement with the tweet (Refer 3(c), Section 5.3.1)
   - (*dm*): Number of users who replied with disagreement to the tweet (Refer 3(c), Section 5.3.1)
   - (*t*): Tweet activity time (Refer 4, Section 5.3.1)

$$Let\ r = (fav + 3 * rt + 1.5 * am + 0.6 * dm)*t \tag{3}$$

Choice of weights.

1. *fav*: A parameter weight of 1 is assigned to *fav*. This weight acts as a baseline for all other weights in the equation.
2. *rt*: A retweet clearly depicts the strongest level of agreement with the media outlets' sentiment. Therefore, due to a stronger level of induced influence than a favorite, it is assigned a weight 3 (>1).
3. *am* : The agreements on a tweet also depict a higher level of influence than a favorite. However, they are less influential than the strongest level of agreement i.e. retweets. Therefore, they are assigned a weight of 1.5.
4. *dm*: Disagreements do not promote the bias shown by the tweet. However, they still increase the reach of the tweet. They are therefore inferred to show influence, but lesser than the baseline i.e. *fav* leading to a weight assignment of 0.6.

Note: The choice of these weights is made in a best effort manner due to the algorithm originating from an unsupervised approach. Proposing an automated technique to obtain these constants is left as future exercise.

The operator * is used to weigh the calculated value with the tweet activity time (*t*). The tweet with a longer activity duration implies that the tweet elicited a good response from the target audience, hence increasing the short term impact of the news bias.

Now let *m* be the immediate mentions and *f* be the number of induced followers. The operator used (*) best captures the reach. Therefore,

$$R = (r + m)*(f + F) \tag{4}$$

4. **Persistence (*P*) of a media outlet:** Denotes the resolute of the media outlet. It captures how consistent a media outlet is about its opinion. The persistence of a media outlet is measured by observing its previous tweets and their polarity scores in relation to the latest tweet released by the media outlet, i.e. *P* is a function of the polarity of past tweets and the latest tweet of the media outlet.

To measure the prior polarity of a media outlet, the exponential moving average of past polarity scores is calculated. The *EMA* for the *PS* series is obtained recursively using Equation (1), where the value of α is chosen as 0.3. Thus,

$$p(t) = 0.3*PS(t) + 0.7*p(t-1) \tag{5}$$

Choice of weights.

The value of a is chosen to be 0.3 to provide fair weightage to both the last tweet and the series of previously made tweets. It was observed that choosing values of $\alpha > 0.3$ masked the effect of the historical tweets and amplified the effect of the latest one. Similarly, values of $\alpha < 0.3$ undermined the latest tweet's contribution to persistence.

Note: The choice of this weight is made in a best effort manner due to the algorithm originating from an unsupervised approach. Proposing an automated technique to obtain this constant is left as future exercise.

The persistence of a media outlet is measured in terms of the similarity of *PS* and *p* (the exponential moving average of past *PSs*).

$$P(t) = 1 - abs(p(t-1) - PS(t)) \,/\, (1 + abs(p(t-1))) \tag{6}$$

The above formulation calculates the 1-D distance between $p(t-1)$ i.e. the value of sentiment so far and *PS*, the current polarity of the media outlet.

The denominator $(1 + abs(p(t-1)))$ is used to measure the range in which the distance between the points is calculated. This is captured by measuring the maximum possible distances to the left (−1) or to the right (1) that *p* could have moved. For example, if $p = 1$ and $PS = -1$, the absolute difference between the two is 2, but this difference is over the range −1 to 1 i.e. 2, thus a persistence of 1-1 = 0. Also, for another example, if $p = -0.5$ and $PS = -1$, the absolute difference between the two is 0.5, but this difference is over the range max (−0.5 to 1 or -1 to −0.5) i.e. 1.5, thus a persistence of 1-1/3 = 0.66.

The pseudocode for determining the persistence is shown in Fig. 7 below:

5.3.3. Bias short term impact score

The Bias Short term Impact Score (*BSIS*) of a media outlet is defined as:

$$BSIS = R*PS*(1 + P) \tag{7}$$

where, *R* is the reach of the tweet, *P* is the persistence of the media outlet

```
-Fetch all the tweets of a media outlet for period to be
analysed
-Initialize p,P,PS arrays with size as no. of tweets and
 all elements initialized to 0
-For each tweet
-       Extract sentiment score of tweet
-       Store sentiment score in corres. PS array slot

-For each tweet
-       Calculate EMA for past tweets and store values
        in corres. p array slot by using formula
        p[i]=0.3*PS[i]+0.7*p[i-1]

-For each tweet
-       Calculate persistence of user after each tweet and
        store in corres. P array slot by using formula
        P[i]=1-abs(p[i-1]-PS[i])/(1+abs(p[i-1]))

-P array contains persistence value of media outlet
after each tweet
```

Fig. 7. Pseudocode for calculating persistence of media outlet after each tweet.

and *PS* the polarity score of the tweet. Note that the constant 1 is added to *P* in order to account for the value of P = 0 for zero persistence.

A bias-short term impact timeline is constructed for each media outlet by plotting the value of BIS for each tweet against time. Each media outlets' BIS-time plot is plotted separately for its sentiment towards different subject sets like the Bhartiya Janta Party (BJP), the Indian National Congress (INC) and the Aam Aadmi Party (AAP).

A BIS timeline is also constructed for a neutral media outlet, which serves as a null model. This study chooses DD-News to serve as the neutral outlet. DD-News is a government-funded media outlet with a reputation of providing the most objective news to viewers. This justifies its selection as the null model to calculate the threshold value for checking bias.

The maximum value of the absolute *BSIS* values on the BSIS-time plot of the media outlet (DD in our case) as obtained from the graph is noted as Neutral Limit (*NL*). The neutral outlet is said to tweet within the polarity threshold of *NL*. It is used as a parameter for calculating the polarity threshold value for a media outlet from its BSIS plot.

The threshold value (T) for a specific media outlet and its BSIS timeline is obtained by –

$$T = NL^{*}IF_{media-outlet}/IF_{DD} \tag{8}$$

Equation (8) given above is used to normalize the value of Neutral Limit *NL* by accounting for the *IF* values of the media outlet under study and the neutral outlet (DD in our case). This, in turn, normalizes the value of *T* which makes it fit for comparison with the *BSIS* for a media outlet.

For example: If a media outlet has a higher value of *BSIS* due to larger values of *F*, *m*, it will still be comparable against a larger value of threshold *T* generated by Equation (8). Equation (8) will generate a comparatively higher value of *T* for the media outlet as the ratio of $IF_{media-outlet}/IF_{DD}$ will be higher for the specific media outlet (due to $IF_{media-outlet}$ being high).

The threshold value plotted on the *BSIS* timeline is used to check for alarming biases.

Any tweet with *BSIS* value above *T* is labelled as 'alarmingly' biased. This bias is 'worrying' or 'disturbing' and triggers a report event. The alarming bias is then said to be identified.

5.3.4. Illustration

Example 5: Consider the tweet "Live | Delhi: Tea stall owner in RK Puram accepts online payments to help customers. #Demonetization" by a media outlet.

Step 1: Calculation of various parameters

F = 1619292.
M = 5046
rt = 12
fav = 40
am = 1
dm = 0
t = 1.4
f = 73879226
m = 100.
PS = 0.4939.

Step 2: Calculation of Influence Factor (*IF*)

From (2).
IF = 8154754512.

Step 3: Calculation of Reach (R)

From (3),
r = 108.5.
From (4).
R = 15741441003.0.

Step 4: Calculation of Persistence (P)

From (5) and (6).
P = 0.448841795403.

Step 5: Calculation of Bias Short term Impact Score (*BSIS*)

From (7).
BIS = 11264306990.9.

Step 6: Calculation of Threshold (T)

IF (for DD-news) = 7212972261.
Threshold value (for DD-News) = 7285358616.66.
IF (for media outlet) = 8154754512.
From (8).
Threshold (for media outlet) = 8170947431.99.

It is evident that the BIS score obtained as *11264306990.9* is greater than the threshold for the media outlet which is *8170947431.99*. Therefore, a report event is triggered and alarming bias is detected.

The same is illustrated using the *BSIS* -timeline as shown in Fig. 8. The red (Fig. 8(a)), blue (Fig. 8(b)) and green (Fig. 8(c))lines respectively are used to draw the *BIS* plots for a media outlet for biases for/against the BJP, INC and AAP. The black line represents the threshold limit. The media outlet can be noticed to be 'alarmingly' biased towards BJP at several instances (the red line crosses the threshold line several times). The plot also shows that the outlet has a slight negative 'alarming' bias towards AAP (the green line in the plot crosses the threshold sometimes). The plot for INC (blue line) stays reasonably within the threshold limit and the media outlets' bias is not labelled as 'alarming' in the given time interval.

5.4. Extending the proposed model

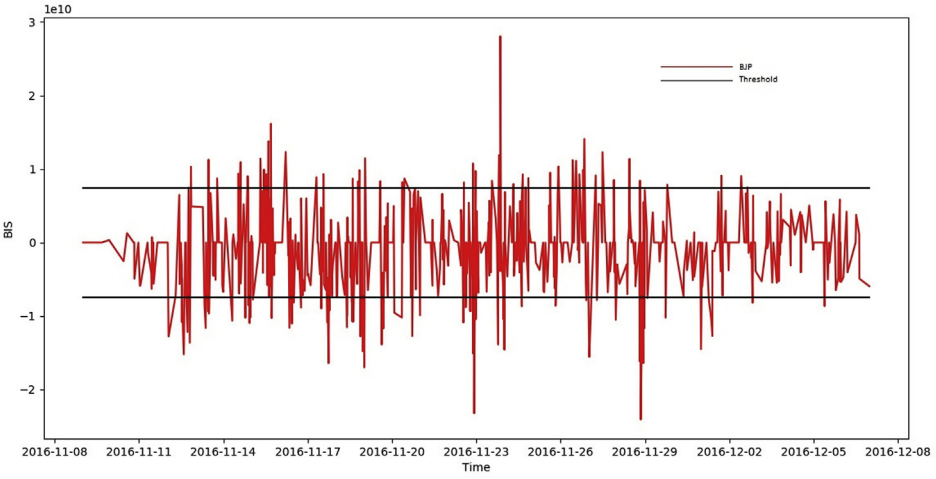
Just like Twitter, sites like Facebook and Reddit offer a continuous stream of personal opinions on the most varied range of topics. The proposed model can be extended for these platforms.

The observations made also reveal that most media outlets are far more biased on live TV than in print and social media. Bias detection in Television media could therefore be an important extension to the model. A naive solution is to convert Live TV stream to Natural Language in text and then apply a similar model to detect biases. Also, media biases in terms of panel representation, think-tank citations and selective news coverage should be dealt with.

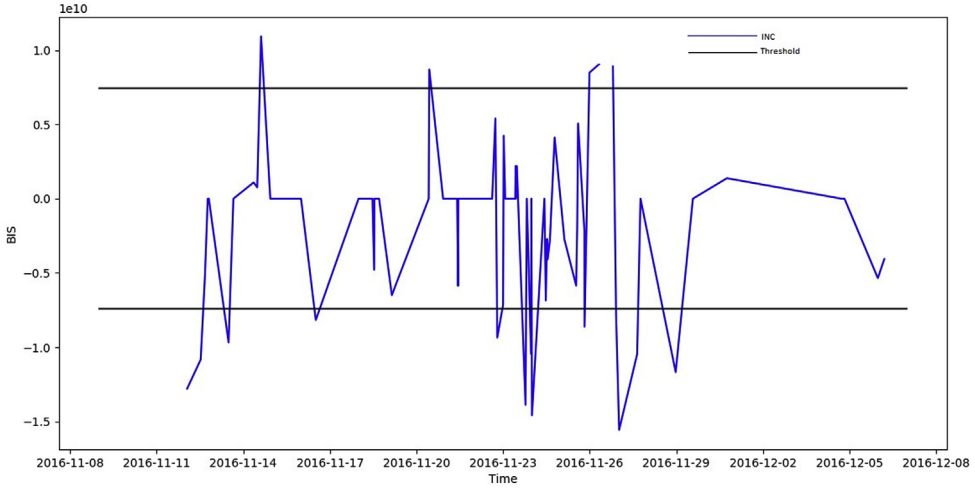
6. Conclusion

As seen by means of a series of experiments, journalists and media outlets are responsible for an opinionated mechanism of news reporting which has a deleterious impact on its consumers. More than 40% of the tweets analyzed were found out to be polar. Also, a fraction of as high as 41% of reactors were conditioned through such 'biased' polar tweets. This shows that an alarming number of people get conditioned by polar, subjective and biased news every day, something, which goes past the naked eye of humans around the world.

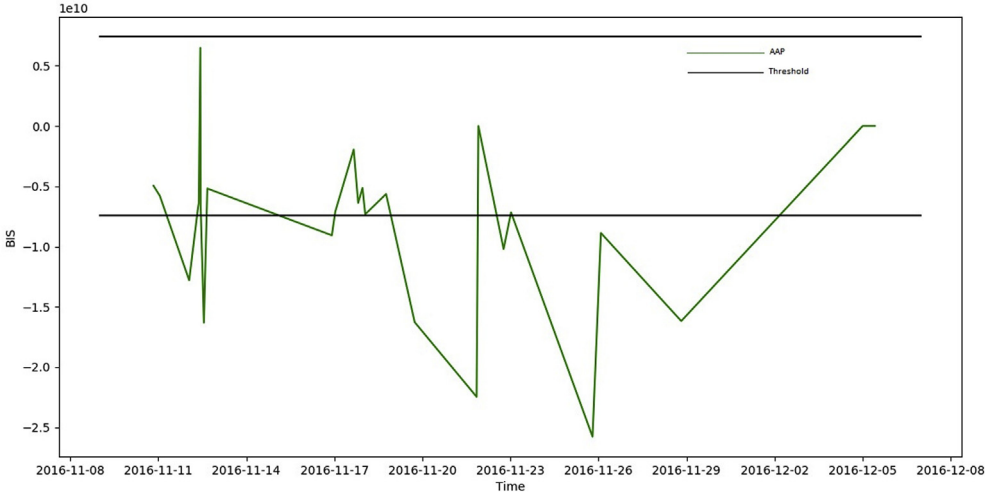
In today's world where digitization has made people's life extremely easy, care must be taken to avoid misuse of the dependence on this medium. Consumers should not blindly trust media sources and use their rationale before developing an opinion on an issue. This research aims to help users detect bias in news media by proposing a model for alarming bias detection. The study also shows how media conditions the opinions



a *BSIS* plots of a media outlet for political party: BJP



b *BSIS* plots of a media outlet for political party: INC



c *BSIS* plots of a media outlet for political party: AAP

Fig. 8. a *BSIS* plots of a media outlet for political party: BJP. *BSIS* plots of a media outlet for political party: INC. **8c** *BSIS* plots of a media outlet for political party: AAP. The above-mentioned process can be represented using the flowchart given below in Fig. 9.

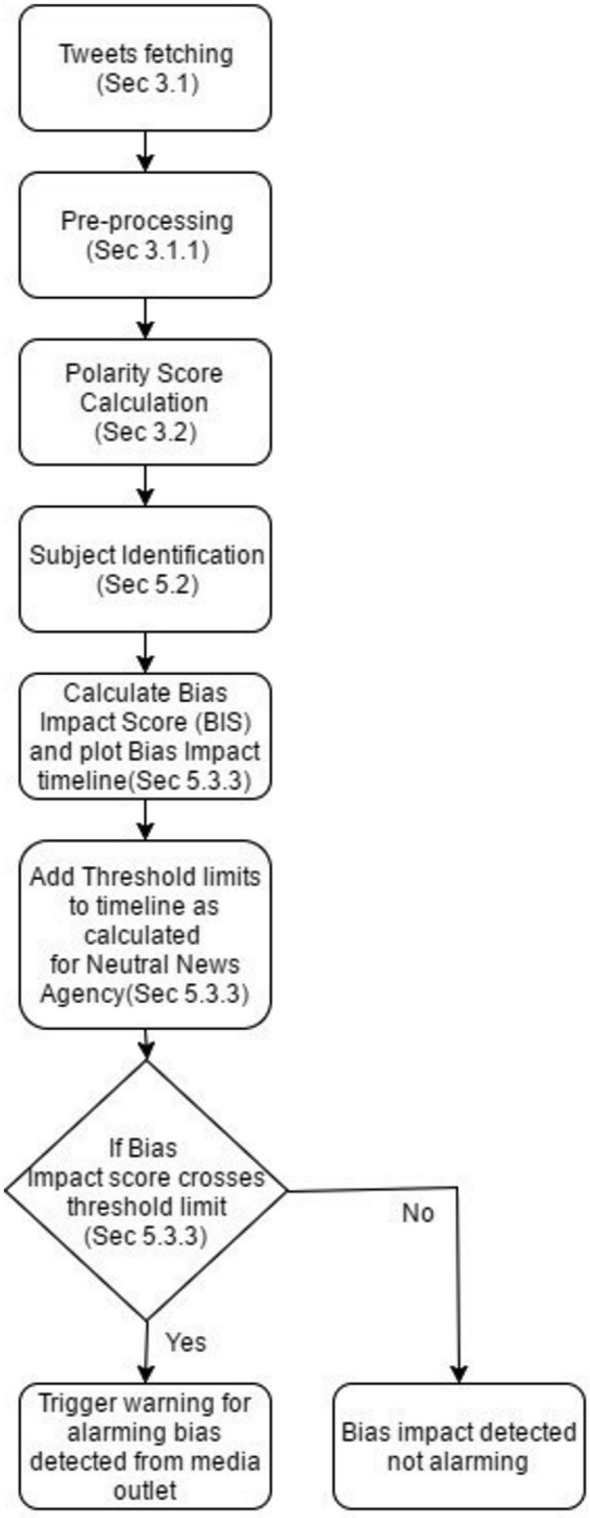


Fig. 9. Flow diagram for checking bias.

of its news-consumers. This justifies the need for a platform to provide safe and healthy balanced news to the news-consumers so that they are not misguided and conditioned by biased media outlets.

In parallel, work needs to be carried out to detect fake news outlets. Fake news is a biased report that is factually incorrect. It is a deliberate attempt to condition the news-consumers in making them believing a falsified report. Tech giants like Google and Facebook have recently started working in this direction. This research aims to address a small fragment of this larger problem of subjective news and its short term impact on the society in large.

Contributors

Swati Aggarwal designed and supervised the study. Siddharth Shikhar was responsible for the collection of data and wrote the manuscript. Tushar Sinha pre-processed and organised the data as well as prepared the graphs and tables used in the study. Yash Kukreti worked on the experiments and Analysis of the study and developed the database. All the authors reviewed the final manuscript.

Declarations

Availability of data

The dataset, as mentioned previously, can be extracted for any generic use-case/topic using the Twitter API [16]. The specific dataset that supports the findings of this study can be procured from the corresponding author upon reasonable request.

Compliance with ethical standards

The authors declare that they have no conflict of interest.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Media Bias-Wikipedia. Accessed 10/06/17, https://en.wikipedia.org/wiki/Media_bias.
[2] Liu B. Sentiment analysis and subjectivity. In: Handbook of Natural Language processing,. second ed. Chapman and Hall/CRC; 2010. p. 627–66.
[3] Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval 2008;2(1–2):1–135.
[4] Hutto CJ, Gilbert E. May). Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth international AAAI conference on weblogs and social media; 2014.
[5] Garon L. A case study of functional subjectivity in media coverage: the Gulf War on TV. Can J Commun 1996;21(3).
[6] White PR. Media objectivity and the rhetoric of news story structure. Language in Performance 2000:379–97.
[7] Lex E, Juffinger A, Granitzer M. June). Objectivity classification in online media. In: Proceedings of the 21st ACM conference on Hypertext and hypermedia. ACM; 2010. p. 293–4.
[8] O'Connell M. Is Irish public opinion towards crime distorted by media bias? Eur J Commun 1999;14(2):191–212.
[9] DellaVigna S, Kaplan E. The Fox News effect: media bias and voting. Q J Econ 2007; 122(3):1187–234.
[10] Gerber AS, Karlan D, Bergan D. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. Am Econ J Appl Econ 2009;1(2):35–52.
[11] Chiang CF, Knight B. Media bias and influence: evidence from newspaper endorsements. Rev Econ Stud 2011:rdq037.
[12] Groseclose T, Milyo J. A measure of media bias. Q J Econ 2005;120(4):1191–237.
[13] Lin YR, Bagrow JP, Lazer D. More voices than ever? quantifying media bias in networks. 2011. arXiv preprint arXiv:1111.1227.
[14] An J, Cha M, Gummadi KP, Crowcroft J, Quercia D. Visualizing media bias through Twitter. In: Sixth international AAAI conference on weblogs and social media; 2012, May.
[15] Dunham WR. Framing the right suspects: measuring media bias. J Media Econ 2013;26(3):122–47.
[16] Twitter Api. Accessed 10/06/17, https://dev.twitter.com/rest/public.