FakeNewsIndia: A Benchmark Dataset of Fake News Incidents in India, Collection Methodology and Impact Assessment in Social Media

Apoorva Dhawan[a], Malvika Bhalla[a], Deeksha Arora[a], Rishabh Kaushal*[a], Ponnurangam Kumaraguru[b]

[a]*Department of Information Technology, Indira Gandhi Delhi Technical University for Women*
[b]*International Institute of Information Technology, Hyderabad*

Abstract

Online Social Media platforms (OSMs) have become an essential source of information. The high speed at which OSM users submit data makes moderation extremely hard. Consequently, besides offering online networking to users, the OSMs have also become carriers for spreading fake news. Knowingly or unknowingly, users circulate fake news on OSMs, adversely affecting an individual's offline activity. To counter fake news, several dedicated websites (referred to as fact-checkers) have sprung up whose sole purpose is to identify and report fake news incidents. There are well-known datasets of fake news; however, not much work has been done regarding credible datasets of fake news in India. Therefore, we design an automated data collection pipeline to collect fake incidents reported by fact-checkers in this work. We gather 4,803 fake news incidents from June 2016 to December 2019 reported by six popular fact-checking websites in India and make this dataset (FakeNewsIndia) available to the research community. We find 5,031 tweets on Twitter and 866 videos on YouTube mentioned in these 4,803 fake news incidents. Further, we evaluate the impact of fake new incidents on the two prominent OSM platforms, namely, Twitter and YouTube. We use popularity metrics based on engagement rate and likes ratio to measure impact and

Email addresses: `adhawan13@gmail.com` (Apoorva Dhawan), `malvikabhalla99@gmail.com` (Malvika Bhalla), `deekshaarora397@gmail.com` (Deeksha Arora), `rishabhkaushal@igdtuw.ac.in (Corresponding Author )` (Rishabh Kaushal*), `pk.guru@iiit.ac.in` (Ponnurangam Kumaraguru)

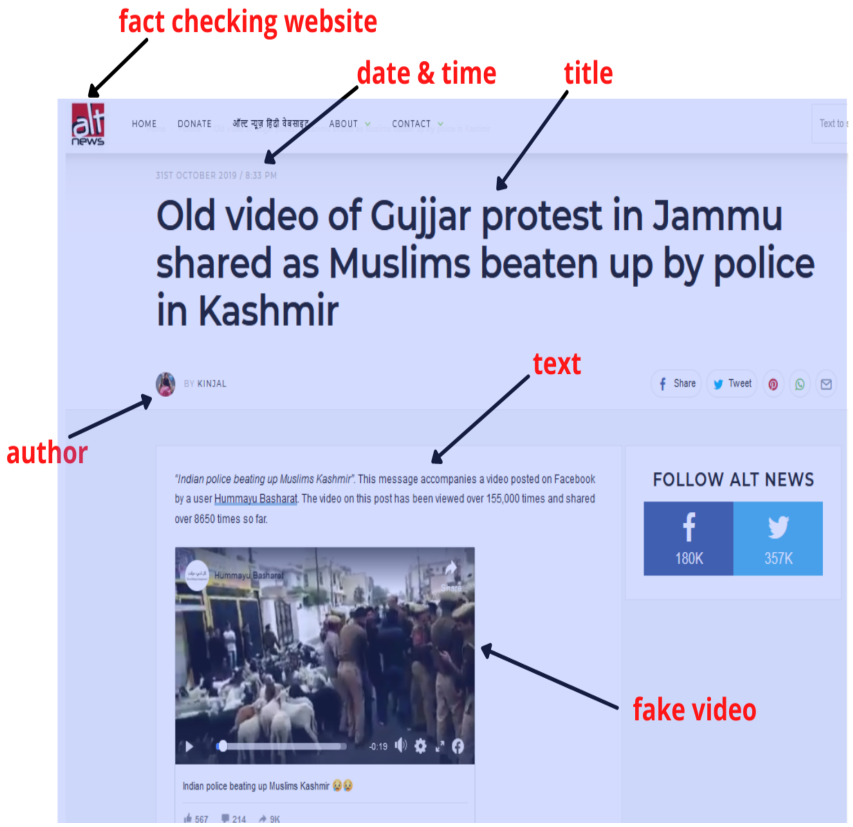
categorize impact into three levels - low, medium, and high. Our learning models use features extracted from text, images, and videos present in the fake news incident articles written by fact-checking websites. Experiments show that we can predict the impact (popularity) of videos (appearing on fake news incident articles) on YouTube more accurately (with baseline accuracy ranging from 86% to 92%) as compared to the impact (popularity) of tweets on Twitter (with baseline accuracy of 37% to 41%). We need to build more intelligent models that predict tweets' impact, appearing in fact-checking incident articles on Twitter as future work.

Keywords: Fake News, Machine Learning, Deep Learning, Social Media, Fact Checking.

1. Introduction

In today's society, the Internet in general and Online Social Media platforms (OSMs) in particular have become all-pervasive. OSMs allow people to interact and access news and information. The Internet serves as a medium to host OSMs websites where creation, sharing, and spreading of information happen at high speeds [1]. With the advancement in communication technologies, information is easily accessible to all. An enormous amount of information is published daily on OSMs. However, the credibility of information [2, 3] on OSMs is a big concern. It is not a trivial task to tell whether the information in circulation is true or false. It requires an in-depth investigation and analysis of the information being shared. It includes, (i) checking the facts, referred to as *fact checking* [4, 5], (ii) assessing the supporting sources, (iii) finding the source of information, and (iv) checking the credibility of authors. Given the scale at which users submit information, moderation of content and fact-checking becomes challenging [6]. It allows malicious users to exploit OSMs to spread fake news. Fake news [7, 8] can be defined as fabricated information created with an intent to cause damage to an individual or organization or to mislead people.

Unfortunately, OSMs have been increasingly used for rapidly spreading false news [9, 10, 11]. There are many motivations for this behavior. Fake News, like clickbait [12], has an element of sensation [13] that draws more readership and engagement. The element of sensationalism in fake news draws more audiences towards it. Moreover, across most OSMs, higher engagement and visibility mean a higher probability of advertisement revenues,



31ST OCTOBER 2019 / 8:33 PM

Old video of Gujjar protest in Jammu shared as Muslims beaten up by police in Kashmir

BY KINJAL

"Indian police beating up Muslims Kashmir". This message accompanies a video posted on Facebook by a user Hummayu Basharat. The video on this post has been viewed over 155,000 times and shared over 8650 times so far.

Indian police beating up Muslims Kashmir 😢😢

Figure 1: An example of a fake news incident reported by the fact-checking website (alt-news). Attributes depicted are title, author, date, time, text, and link to the fake video. The fake news incident explains that an old video was re-used in a wrong context, in order to falsely push a narrative.

so spreading fake news is financially more rewarding for users who create and spread it [14]. However, fake news is undesirable; it has adverse effects on society [15]. It hampers the credibility of an individual or organization. It plays with people's emotions; for instance, it can make people feel happy or sad or scared. It is even responsible for physical harm; for instance, in one case, the news shared via WhatsApp led to murder, and at least 31 people were killed in 2017 and 2018 due to mob attacks fuelled by rumors on WhatsApp.[1] Considering the negative impacts that fake news has on society, we frame the following objectives for this work:

1. To design a data collection methodology for automated fake news collection in the Indian context.
2. To quantitatively perform an impact assessment of fake news circulating within India on OSMs.

Numerous prior works in the area of fake news have addressed the issue by curating datasets like BuzzFeedNews [16], LIAR [17], and CREDBANK [18] to name a few. However, to the best of our knowledge, no credible dataset exists for the Indian context that could help fight fake news in India. In the fake news research landscape, India poses unique and unprecedented challenges; India is one of the most diverse civilizations with many religions, languages, and political parties, with the second largest population in the world. Bounced with technological advancement, the smartphone penetration is high [19], so the possibility of fast spread of fake news is quite high. Due to the digital divide, there is a vast population who is not tech-savy and also not very literate, and they fall into the trap of believing in fake news [20]. Given these challenges in India, it is imperative that a dataset explicitly focusing on the Indian context be created. To this end, in the first part of the work, we design a data collection approach to collect the fake news incidents reported by the six most popular fact-checking websites in India, referred to as **FakeNewsIndia**. The first version [2] of FakeNewsIndia dataset comprises 4,803 fake news incidents happening in India , collected during June 2016 to December 2019. We plan to update this dataset periodically using our automated data collection scripts. We understand that in this first version, we have collected fake news incidents that are being reported in the English language, and it may appear to be a limitation. However, the fake news that we collect is related to the incidents happening in India. We believe that it is an important first step towards collecting fake news incidents within India. These incidents happen across different regions in India, and are reported by both English media fact checkers (which we have captured in our work) and also by regional media fact checkers. Collecting fake news incidents being reported by regional fact checkers in regional languages is not addressed by

[1] https://www.bbc.com/news/world-asia-india-47797151

[2] https://github.com/rishabhkaushal/fakenewsincidentsIndia

our work. However, we believe that the fake news incidents captured by us also have visibility and impact among different regional viewers.

In Figure 1, we depict an example of a fake news incident[3] reported by the fact-checking website (alt-news) along with the attributes, namely, title, author, date, time, text, and link to the fake video.

In the second part of our work, we extract the social media links mentioned in these fake news incidents pointing towards two OSMs: YouTube and Twitter. Most of these links point to either text or image or video or a combination of these. In total, it turned out that 4,803 fake news incidents have 5,031 links to tweets on Twitter and 866 unique video links on YouTube. Using the features extracted from the text and the image of the fake news posts on OSMs, we build machine learning models to predict the impact of these fake news posts. As a metric to assess the impact, we use popularity metrics based on engagement rate and likes ratio [21]. We categorize the impact into three classes (referred to as *impact levels*), namely, low, medium, and high. Low level impact means that not many users on social media platforms engaged or liked the fake content. Impact of medium level means that only moderate proportion of users engaged or liked the fake content. Whereas a high impact level means that large number of users engaged or liked the fake content. Our model helps in answering a critical question *'From the text and image of fake news posts present on OSMs, can we predict the impact (engagement) measured in terms of their popularity?'* If we can successfully make this prediction, we will prioritize our response towards the fake news incident. We can address those fake news incidents, which we expect would draw more attention at a top priority. Our experiments show that we can accurately predict the three impact levels with an accuracy range of 86% to 92% for the fake videos on YouTube, which is quite encouraging. However, predicting the three impact levels for tweets appearing in fake news incident articles turns out to be in the range of 37% to 41%, which is near to random prediction. We shall explain the reasons for these results, which opens the path for more work in this direction. In a nutshell, from this work, we conclude that predicting the impact of fake news incidents on YouTube is quite accurate, whereas predicting impact on Twitter is quite tricky. To summarize, we make the following contributions.

[3] Link:https://www.altnews.in/old-video-of-gujjar-protest-in-jammu-shared-as-muslims-beaten-up-by-police-in-kashmir/

- We gather 4,803 fake news incidents using six popular fact-checking websites in India, containing 5,031 links to Twitter and 866 links to videos on YouTube. We refer to this dataset as **FakeNewsIndia**, and make it available to the research community here.[4]

- We perform impact assessment of fake news incidents on YouTube and Twitter. We use likes ratio and engagement count as the metrics for measuring impact. We obtain an accuracy range of 92% - 86% for impact prediction of fake video posts on YouTube. However, on Twitter, the accuracy range is 41% - 37%, using text and image-based features in the tweets appearing in fake news incident articles.

One possible explanation for better results on YouTube than Twitter is as follows. As per GlobalStats[5], for the month of October 2021, the market share of YouTube and Twitter in India is 16.69% 5.78%, respectively. This would imply more user interactions (views, likes, engagements) on YouTube than on Twitter. Consequently, the availability of user interactions data needed to build a good machine learning based prediction model is more likely for YouTube than Twitter. Nevertheless, we believe that this work is helpful to access any new fake news in terms of its impact, and we can prioritize our response accordingly. Moving further, in the next section, we discuss the current datasets available in the domain of fake news detection. After identifying the gaps, we discuss our proposed approach to collect a comprehensive dataset on fake news in the Indian context and a fake news impact assessment methodology.

2. Related Work

In this section, we discuss the prior works in three parts. In the first part, we explain the prominent datasets available in the domain of fake news. In the second part, we discuss the key works related to the fake news detection. Moreover, in the last part, we focus on earlier works that study the impact of fake news on OSMs.

[4] https://github.com/rishabhkaushal/fakenewsincidentsIndia

[5] https://gs.statcounter.com/social-media-stats/all/india

2.1. Fake news datasets

We can get news from different sources like OSMs, search engines, the homepage of news agencies, or fact-checking websites. Determining the credibility of news is not a trivial task. Many researchers have done an in-depth analysis of fake news and determined its veracity. Different studies have led to the creation of publicly available datasets on fake news. We highlight a few prominent ones below:

1. BuzzFeedNews[6]: This dataset [16] has a collection of news items published on Facebook by nine news agencies for a week during the 2016 US elections (19th to 23rd September and 26th-27th September). It has 1,627 articles, of which 826 are mainstream articles, 545 are right-wing articles, and 356 are left-wing articles.
2. LIAR: Wang et al.[17] prepared a dataset comprising of data from fact-checking website - Politifact. Dataset comprises of short statements made by American politicians. Data was collected using the API of the website. It has over 12,836 statements labeled by humans. Each record is assigned a label based on truthfulness: true, mostly-true, barely-true, half-true, pants-fire, and false.
3. BS Detector[7]: It is created by collecting data using the browser extension BS-detector. It checks each link for its veracity against a manually created list of domains.
4. CREDBANK: Mitra et al. [16] created this dataset[8] which comprises 60 million tweets collected for over 96 days from October 2015. All tweets are related to 1049 events that occurred in the real world. It has been annotated by 30 people from the Amazon Mechanical Turk.
5. BuzzFace: Santia et al. [22] created BuzzFace[9] which is an extension of the BuzzFeedNews dataset. It also stores the comments on the fake news stories published on Facebook. The dataset contains 2,263 articles with 1.6 million comments.
6. FacebookHoax: Tacchini et al. [23] created FacebookHoax[10] using the Facebook Graph API. It has non-hoax and hoax scientific news articles. It contains 15,500 posts from 14 hoax pages and 18 non hoax pages.

[6] https://github.com/BuzzFeedNews/2016-10-facebook-fact-check

[7] https://github.com/thiagovas/bs-detector-dataset

[8] http://compsocial.github.io/CREDBANK-data/

[9] https://github.com/gsantia/BuzzFace

[10] https://github.com/gabll/some-like-it-hoax

We observe that all the prominent datasets discussed above for fake news are not specific to India. In the final stages of writing this paper, we found a recent work by Singhal et al. [24] which collected and annotated fake news incidents in India inclusive of regional languages. Instead, we focus on Indian fake news incidents in the English language and perform impact assessments.

2.2. Fake News Detection

There are various definitions and terms that can fit under the umbrella of fake news, namely, disinformation, misinformation [25, 26, 27], hoaxes [28, 29, 30], rumors [31, 32, 33]. In this subsection, we shall discuss some of these essential works with the aim of being indicative and not exhaustive.

Shu et al. [34] presents the most comprehensive review on fake news detection. Accordingly, the detection techniques derive features from social context and fake news content. Linguistic and visual features are extracted from the fake news content using a source of news, headline, body text, and image/video in the news. From the social context, they extracted features from user-level, post-level, and network-level. Next, we discuss some of the recently proposed approaches for the detection of fake news. The work of Monti et al. [35] is based on the premise that the spreading of real and fake news forms different propagation patterns. Leveraging propagation is also advantageous because these patterns are independent of the underlying language. Their model makes use of graph convolution using propagation and network structure-based features. Given that most fake news detection approaches are supervised, Yang et al. [36] investigated whether the unsupervised method can be adopted. They consider users' credibility and truths as random variables, besides leveraging other users' engagement on the news posts. They propose a Bayesian model with the Gibbs sampling approach as the unsupervised solution. Okoro et al. [37] proposed a hybrid approach combining machine-based and human-based intelligence for the detection of fake news. Guo et al. [38, 39] worked on the intuition that fake news invokes a lot of emotions among humans. They proposed a technique called the Emotion-based Fake News Detection framework (EFN), which uses emotion representation vectors as features extracted from the emotions in comments and content. Stahl et al. [40] proposed a model which combines semantic analysis with Support Vector Machines (SVM) and Naive Bayes to detect fake news. Tschiatscheck et al. [41] exploited the signals from the crowd, in other words, the users who flag the news as fake, for building a Bayesian inference based algorithm referred to as DETECTIVE. Lu et al. [42] proposed Graph aware Co-Attention Networks (GCANs), which takes the text of the tweet and users who have retweeted the tweet as input to flag whether the tweet is fake or not, along with explanations. Tacchini et al. [23] classified Facebook posts into non-hoaxes and hoaxes based on users who engaged (liked) with those posts. Shu et al. [43] relied on the social interactions that users make on OSMs for detecting fake news. They performed a detailed understanding on users to categorize them as experienced or otherwise in flagging fake news. Liu et al. [44] proposed a deep neural network for early detection of fake news. They extracted user features and text features from the responses to the fake news and used an attention mechanism that is position-aware and moved over time windows.

Wu et al. [25] defined fake news as news that is incorrect, or information is known to be false. They studied the propagation of fake news and compared different detection methods. Yu et al. [26] introduced a Convolutional Neural Network (CNN) based method called Convolutional Approach for Misinformation Classification (CAMI). They extracted features spread across input sequences and interactions. Jain et al. [27] proposed an approach to automatically detect a rumor based on whether a reliable source has posted it (e.g. news channel) or not.

Qazvinian et al. [31] define a broader term rumor which is referred to a statement whose truth value is either true or false or can not be verified. Fake news is rumors which are known to be false. Takahashi et al. [32] detected rumors on the social media platform Twitter, during a disaster, for example, earthquake. Dayani et al. [33] performed a retrospective analysis of rumors using the dataset proposed by Qazvinian et al. [31]. Zhang et al. [45] introduced an automated method to detect rumors based on shallow and implicit features in a message. Kwon et al. [46] found that temporal and structural features over a more extended period are able to distinguish non-rumor from rumor. Ma et al. [47] proposed neural network-based two recursive models for learning rumor representation and detection.

2.3. Impact of Fake news

This sub-section shall discuss prior works that measure the impact of social media posts in general and fake news posts in particular. Some works [48, 34] focus on identifying factors that affect the popularity of social media posts, whereas other works [49, 50, 21] focus on predicting popularity of social media posts. Swani et al. [48] investigated and found the key factors that help in popularizing a post on social media. Mishra et al. [49] proposed Recurrent Neural Network (RNN) based approach to model and predict the popularity on social media. Xu et al. [50] proposed an approach to forecast popularity of videos. Shu et al. [51] proposed FakeNewsTracker that collected news in an automated manner and created a dashboard to measure the impact of fake news among the collected news by understanding factors that affect fake news. Various works towards calculating the engagement rate of a post on social media platforms have been explored. They are helpful for fake news impact determination. Aldous et al. [21] predicted the audience engagement, precisely the number of likes and comments through this work. Their work consisted of 676,779 posts on social media taken from 53 news outlets collected for 8 months on four platforms (Twitter, Facebook, YouTube, Instagram, and Twitter). The prediction of the audience engagement was based on social media platform factors and linguistic features of the post.

2.4. Research gaps

Based on our study of past works, we find no benchmark dataset for the fake news problem for the Indian context. Moreover, these datasets have their own limitations. For instance, the BuzzFeedNews dataset comprised only headline text for each news piece annotated into left and right wing articles. LIAR dataset was mostly made up of short statements collected from various speakers (mostly political). BS-detector collected unverified links based on browser extension, and is not specifically for fake news. CREDBANK performed a collection of tweets and determined their credibility. We observe that none of these datasets provide a collection of fake news incidents that provides links to social media posts. Instead, our dataset comprises title, author, date, time, name of fact checker, and an entire description containing links to social media posts.

2.5. Objectives

Fake news is always deceptive, and its spread impacts millions of people. Since fake news in India is spreading rapidly, it is imperative to curb this evil. So, our main objective is to collect fake news, and study the impact of fake news, and measure user engagement with it. We have the following objectives:

1. To design a data collection methodology for automated fake news collection in the Indian context. To this end, we curated fake news from

various fact-checking websites in India based on their Alexa rankings. After crawling the websites, we curate samples that contain URLs pointing to fake posts on different OSMs. The dataset we create is a multimodal data set because it has links to both images and video links.

2. Impact of fake news: We quantitatively predict the impact (into three classes namely, high, medium, and low) that the fake news is likely to create on social media. We build models to predict this impact class using various features like the text of the fake news, an image circulated, or the video associated with the fake news. This work would help anyone with n number of fake news incidents to prioritize which fake news one should address first based on the impact class.

3. Fake news data collection

Different fact-checking websites report various fake news incidents after verifying them. In this section, we present the approach that we use for fake news data collection.

Algorithm 1 Data Collection Algorithm

```
procedure GetFakeNewsIncident($fc\_name$)
    $fc\_url$ = get_FactChecker_URL ($fc\_name$)
    $fci\_list$ = get_All_FC_incidents ($fc\_url$)
    for $fc\_article \in fci\_list$ do
        if $fc\_article \in Database\ (FC\_incidents)$ then
            continue
        else
            $Database\ (FC\_incidents) = Database\ (FC\_incidents) \cup fc\_article$
        end if
    end for
end procedure
```

3.1. Collection Methodology

We collect Indian fake news incidents from different fact-checking websites. As depicted in Figure 2 and explained in Algorithm 1, we crawl on

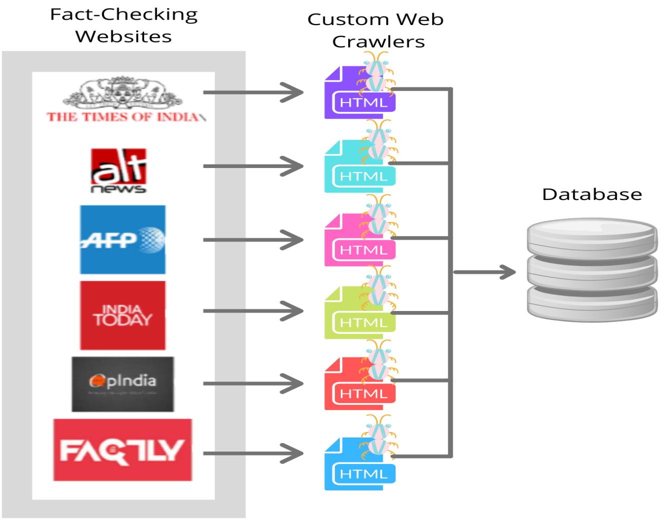


Figure 2: Methodology for fake news data collection. We depict the six fact checkers popularly used to detect fake news in India. We write code to automatically crawl these fact-checkers, collect the requisite data presented in different web page structures, and store in a common database.

these websites to collect new fake news incident reports. When our crawler comes across a particular news incident, it checks whether it is already present in our records or not. We detect duplicate fake news incidents as follows. First, we remove the stop words from the title of fake news incident, and then compare words in the titles, irrespective of the order in which they appear. If common words are present in the titles, then we flag news incidents as duplicate else we consider them distinct. If present, we move on to the next fake news incident. Otherwise, we crawl over that fake news incident, store the collected data in a CSV file, and subsequently into the MongoDB database. We search for links to social media platforms on which fake news was spread from each news incident. We use these links to extract meta-data from various social media APIs. Many websites are present that fact checks a news incident and reports it as fake after many scrutinies. In order to select from which fact-checking websites we have to collect our data, we use Alexa rankings. Alexa[11] ranking is a metric that ranks a website based on

[11] https://blog.alexa.com/marketing-research/alexa-rank/

its popularity. The lower the Alexa rank number, the higher is the customer engagement with that website. In Table 1, we mention the Alexa ranks of popular fact-checking websites in India. Accordingly, we select the following fact-checking websites as they have good Alexa rankings: (1) Times of India, (2) India Today, (3) AFP FactCheck, (4) OpIndia, (5) Alt News, and (6) Factly. All the fact-checking websites have different structures, and they present the fake news incidents in different HTML structures. So to crawl the data from them, we write separate crawlers for each of them. Moreover, in every fact-checking website, the page limit is different. So it is not easy to write a generalized crawler that can apply to all the websites. So, we write a crawler program that starts collecting fake news from a fact-checking website sequentially and stops when all the fake news incidents are collected. We collect the data periodically and store it in the MongoDB database.

Table 1: We depict fact checking websites and their Alexa Rankings as on 27th September 2019

Fact Checking Website	**Alexa Ranking**
Times Of India	148
India Today	606
AFP FactCheck	14,162
OpIndia	22,120
Alt News	64,725
Factly	224,414

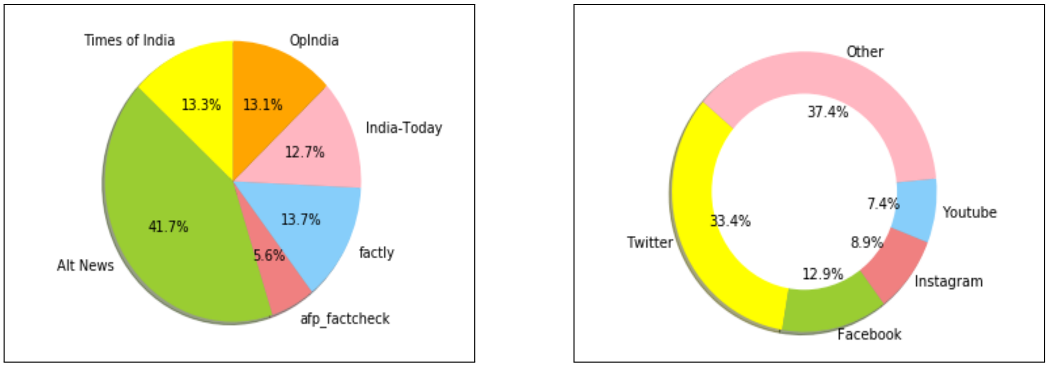
In Table 2, we list down the data attributes associated with each fake news incident that we collect from different fact-checking websites. We faced various difficulties in data collection. This is because the same piece of code for automating the data collection process is not reusable as the structure of every fact-checking website is different. The data we collect from different fact-checking websites has different date formats, and we computationally convert them into the same format so that we can sort the data based on date. Other problems we face are that some of the posts on different social media platforms are deleted, and hence are present in archives in the form of screenshots. Since we collect data from different fact-checking websites, they may capture the same fake news, leading to duplicates in the dataset, which were removed as discussed earlier.

Table 2: List of data attributes collected for each fake news incident. Refer Figure 1 for an example of fake news incident.

Field	Description
fact_checker	Name of the fact checking website.
fake_news_url	Link of the fake news incident.
fake_news_date	Date when the fake news incident is reported.
fake_news_title	Title of the fake news incident.
twitter_tweets	Twitter tweet link used for circulating fake news.
facebook_post	Facebook post link used for circulating fake news.
youtube_link	Youtube video link used for circulating fake news.
instagram_post	Instagram post link used for circulating fake news.
fake_news_imgs	Link to various images used for circulating fake news on other platforms, for example, WhatsApp.
content	Content of the fake news incident.

3.2. Collected Data Analysis

In this sub-section, we perform data analysis on the collected fake news incidents that are reported by fact-checking websites in India. In Figure 3a, we depict the number of fake news incidents reported by fact-checking websites. We observe that Alt news provides the largest collection of news incidents for the duration in which we performed the data collection process. In Figure 3b, we plot the distribution of the presence of social media URLs in the fake news incidents that we collected. We observe that many fake news incidents are discussed and circulated on Twitter and Other platforms (WhatsApp). To understand the number of fake news incidents being reported by fact checking websites, in Figure 4a, we plot the top 5 days on which the maximum number of fake news incidents were reported daily. Our objective is to know maximum number of fake news incidents being reported per day. After analyzing fake news incidents, we find that we usually have 10-15 fake news incidents per day. However, on 4 September 2018, we saw the maximum fake collection of 26 fake news incidents. We may think that



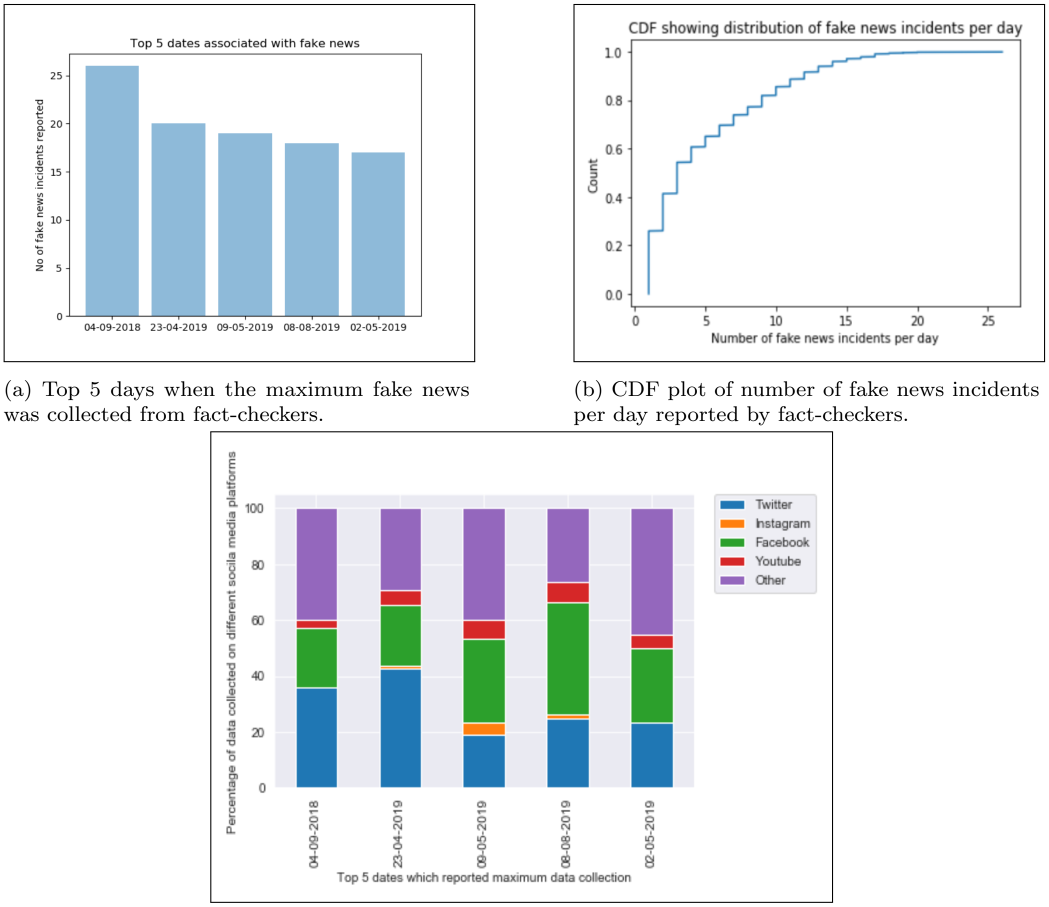
(a) Distribution of data collected from different fact checking websites. AltNews has most of the contribution.

(b) Graph depicting which social media is the major carrier of fake news. Twitter is the most common platform.

Figure 3: Data distributions in our fake news incidents dataset.

the high frequency of fake news incident reporting could be because of political factors (elections, etc) or any other major event. However, it turns out that high frequency is due to many factors. For instance, on 04-09-2018, there are fake news incidents related to floods in Kerala (a state in India), bank holidays, and other political controversies. It is merely a coincidence that total fake news incidents being reported by these different fact checkers, when seen in aggregate, turns out to be higher for a few particular days. In Figure 4b, we draw CDF plot of number of fake news incidents collected per day. On the X-axis, we plot the number of fake news incidents being reported each day, and on the Y-axis we plot the number of days (count) for which that many fake news incidents were reported. For almost 60% of the days, the number of fake news incidents reported by fact checkers were 5 or less. Furthermore, we find that less than 5 fake news incidents are reported by fact checkers for most of the days. In Figure 4c, we plot the distribution of social media URLs that appear in the fake news incidents in the top 5 days. Our aim is to understand the social media platforms which are predominantly involved in the spread of fake news incidents. Across all these days, we observe that Twitter, Other (mainly WhatsApp), and Facebook appear in most of the fake news incidents. Instagram is rarely being used for the spread of fake news incidents. However, YouTube does appear particularly when fake news involved video content.

Next, we perform named entity recognition on the 'content' attribute of a fake news incident. Rahul Gandhi, Narendra Modi, and Sonia Gandhi are



(a) Top 5 days when the maximum fake news was collected from fact-checkers.

(b) CDF plot of number of fake news incidents per day reported by fact-checkers.

(c) Distribution of social media platforms on top 5 dated when maximum fake news was collected.

Figure 4: Distribution of fake news incidents.

prominent personalities present in fake news incidents. All the top personalities are from politics and not from any other domain, so we can conclude that fake news mainly revolves around politics. In other words, politics appears to be the most commonly occurring domain for which fake news are produced and consumed. Other commonly occurring entities are Indians, Hindus, and Muslims in the fake news incidents. BJP, Congress, and organizations like Google, Facebook find a place in fake news incidents. Delhi and West Bengal are the major states that we found in fake news incidents. World cup, new year, Olympics are some of the everyday events that find their place in fake news.

4. Impact Assessment Methodology

This section explains our methodology for impact assessment of fake news incidents on social media platforms. For impact measurement, we rely on engagement metrics to compute the impact scores. Recall that URLs to Twitter were most frequently found in fake news incidents, followed by URLs to Facebook and Youtube. We choose Twitter and Youtube as the two social media platforms for an impact assessment on collected data because a substantive number of fake news incidents had URLs pointing to posts on these platforms. We could not work on Facebook because the platform does not allow the accessible collection of engagement data through Facebook API. We extract the following features from the YouTube video title, description, video, duration, and category for YouTube video. For tweets on Twitter, we extract text-based features from tweet text and image-based features from images embedded in the tweet.

4.1. Engagement & Impact Score

The impact of a social media post is directly proportional to the engagement received and popularity of the post. Users engage on a social media platform by sharing and liking the posted content. For impact assessment of a video on YouTube, we collect view counts, like count, dislike count, comment count, and subscriber count. For impact assessment of a tweet on Twitter, we obtain favorites count, retweet count, followers & followee count of the user who posted the tweet, replies count, and listed count. Inspired from the work of Aldous et al. [21], we use the following formulation to measure impact in terms of engagement rate and likes ratio.

YouTube: The engagement rate ($ER_{YouTube}$)is the ratio of the sum of the view count (v), like count (l), and comment count (c) on a video to the sum of dislike count (d) and subscriber count (s) for the channel who posted the video. The likes ratio ($LR_{YouTube}$) determines the extent to which a video was liked and promoted by people. Here, f denotes the number of likes received by the video, dy represents the number of days from the posting day till the day we collected video statistics, and m refers to the maximum likes received among all the videos in the dataset.

$$ER_{YouTube} = \frac{v + l + c}{d + s} \tag{1}$$

$$LR_{YouTube} = -(\log \frac{f + 1}{dy + m}) \tag{2}$$

Twitter: The Engagement Rate ($ER_{Twitter}$) is the ratio[12] of the sum of the retweets (r) of the tweet and the favorite (fav) count for a tweet to the sum of the followers (fl) of the tweet creator, listed count (l) on a tweet and the friends (fr) of the user who posted the tweet. Likes ratio ($LR_{Twitter}$) is defined as the negative log of the ratio of the sum of the number of likes (l) on a post to the sum of the number of days (dy) since the tweet was posted and the maximum number of likes (m) received among all the tweets in our dataset.

$$ER_{Twitter} = \frac{r + fav}{l + fl + fr} \tag{3}$$

$$LR_{Twitter} = -(\log \frac{l + 1}{dy + m}) \tag{4}$$

4.2. Features Extraction

We extract features from the text and images associated with the URLs found in fake news incidents, which points to tweet on Twitter and videos on YouTube. We extract these features with the help of different APIs. For YouTube, we explain in Table 3 all the features extracted from video title text, description text, video duration, and category. We use YouTube Data API to obtain the video title, category, and duration. For obtaining detailed sentiment scores, we employ Amazon Comprehend API on the text of the video title. We also obtain the count of the number of dates, items, locations, events, organization, persons, quantities, titles, and other entities present in the video title and description text with a confidence score $\geq$ 0.7, obtained using Amazon Comprehend API. Similarly, for Twitter, in Table 4, we explain all the text and image features extracted from tweet text and image in the tweet, respectively. We use Twitter API to obtain the text of the tweet. We employ Amazon Comprehend API to extract detailed sentiment and entities (count of the number of dates, items, locations, events, organization, persons, quantities, and titles) present in the tweet text. In addition, we also use Sight Engine API to obtain nudity types (raw, partial, and safe); scores for a weapon, alcohol, and drugs detected in image in the tweet; sharpness, brightness, contrast, r_value, g_value, and b_value; and the probability of artificial and natural text present in the image in the

[12] www.socialbakers.com/blog/467-formulas-revealed-the-facebook-and-twitter-engagement-rate

Table 3: Table showing various features extracted for Youtube impact model generation.

Names of Features	Descriptions
ytitle_length	Length of video title text, obtained using YouTube Data API.
mixed_score, pos_score, neut_score, neg_score	Represents likelihood of mixed, positive, neutral, and negative sentiment in video title text, obtained using Amazon Comprehend API.
tot_date, tot_item, tot_loc, tot_event, tot_org, tot_person, tot_quantity, tot_title, tot_other	Represents the count of the number of dates, items, locations, events, organization, persons, quantities, titles, and other entities encountered in a video title and description text with a confidence score ≥ 0.7, obtained using Amazon Comprehend API.
category, duration	Denotes the category and duration of video, obtained using YouTube Data API.

tweet. Next, we present some insights from the analysis of the feature values. From the study of sentiments, we observe that mostly the fake news incident that spreads on the Twitter platform is neutral. The most commonly found emotion in images circulated with fake news incidents is *calm*, and most images do not have any emotions. On analyzing the distribution of category values, we find that videos of the categories 'News and Politics' and 'People and Blogs' are the most commonly found fake news incidents.

4.3. Model Construction

In this sub-section, we explain different settings for constructing machine learning-based models. The underlying objective of all models is to predict impact scores (explained in Sec 4.1) measured by either Likes Ratio (LR) or Engagement Rate (ER) of fake news incident. The features we compute (explained in Sec 4.2) are passed as input to these models are derived from the video, text (tweet), or image related to the fake news incident that appears on two social networks, namely, YouTube and Twitter. As depicted in Figure 5, we derive explanatory variables (features) from the URLs appearing in fake news incidents that point to YouTube and Twitter. From YouTube, we extract features from the video title and description text. In the case of

Table 4: Table showing various features extracted for Twitter impact model generation

Names of Features	Descriptions
ytitle_length	Length of tweet text, obtained using Twitter API.
mixed_score, pos_score, neut_score, neg_score	Represents likelihood of mixed, positive, neutral, and negative sentiment in tweet text, obtained using Amazon Comprehend API.
tot_date, tot_item, tot_loc, tot_event, tot_org, tot_person, tot_quantity, tot_title, tot_other	Represents the count of the number of dates, items, locations, events, organization, persons, quantities, titles, and other entities encountered in tweet text with a confidence score ≥ 0.7, obtained using Amazon Comprehend API.
emotion_type, emotion_score	Denotes the type and confidence value of emotion of image in tweet, obtained using Amazon Rekognition API.
raw, partial, safe	Score for raw, partial, and safe nudity detected in image in tweet, obtained using Sight Engine API.
weapon, alcohol, drugs	Score for weapon, alcohol, and drugs detected in image in tweet, obtained using Sight Engine API.
scam_prob	Probability whether a scammer is identified in the image in tweet, obtained using Sight Engine API.
sharpness, brightness, contrast, r_value, g_value, b_value	Denotes sharpness, brightness, contrast, r_value, g_value, and b_value in the image in tweet, obtained using Sight Engine API.
artificial_text_prob, natural_text_prob	Denotes the probability of artificial and natural text present in the image in the tweet, obtained using Sight Engine API.

Twitter, we extract features from the tweet text and image (if any) appearing

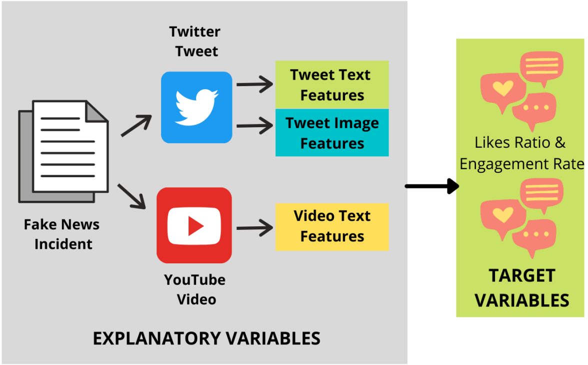


Figure 5: Experiment Settings for Prediction Models of Impact Assessment.

in the tweet.

5. Evaluation and Results

In this section, we explain the various evaluation strategies adopted and their results.

5.1. Impact Assessment on YouTube

In the first part, we evaluate the performance of models that use features derived from YouTube video title, description, and meta-data to predict likes ratio ($LR_{YouTube}$) and engagement rate ($ER_{YouTube}$). Specifically, we ask two hypotheses as below:

- H1: Can YouTube video features based on the video title, description, and meta-data of fake news incident predict **likes ratio** received on that video?
- H2: Can YouTube video features based on the video title, description, and meta-data of fake news incident predict **engagement rate** received on that video?

To answer H1, we construct four linear regression models (M_1, M_2, M_3, M_4) using different features (as depicted in Table 5) to predict likes ratio ($LR_{YouTube}$). The three models uses the three sets of features and the fourth model uses all the features combined together. In each model, we split the

Table 5: Regression results of prediction of likes ratio ($LR_{YouTube}$) of YouTube videos appearing in fake news incidents.

Features Used	**RMSE**
Sentiment features (mixed_score, positive_score, low_score, neutral_score) of text in YouTube video titles	4.503
Features based on entity count (like organization, events, etc) present in the text of YouTube video titles	4.478
Duration & length of YouTube video title	4.406
Combined features	4.396

data into 80% train and 20% test, and perform hyper-parameter tuning to get the best results. The standard deviation (SD) in $LR_{YouTube}$ (target variable) is 4.17 with 3.83 and 22.28 as the minimum and maximum values, respectively. We employ evaluation metrics as RMSE and R-square value. As evident from Table 5, we observe an average RMSE of 4.446 which means that we are unable to predict likes-ratio with certainty. The reason we consider the RMSE values in Table 5 as high is also reaffirmed from Table 6 where we present the range of likes ratio. All the regression models have high RMSE, and consequently, they have negative R-square values. Given the poor performance, we conclude that it is challenging to predict the likes ratio (which indicates the number of likes received) of YouTube videos that appear in fake news incidents using the video title, description, and metadata. Since regression did not give promising results, we turned our attention to the question, whether we can predict *category* of likes ratio. To perform this classification experiment, we need to convert the likes ratio from a real number to a category. We perform K-S test to find whether likes ratio distribution follows normal distribution. We obtain test statistic of 0.997 with p-value less than 0.5, which suggests that null hypothesis can be rejected. In other words, the likes ratio distribution does not follow normal distribution. In Figure 6, we depict distribution of Likes Ratio of videos on YouTube appearing in fake news incidents. Subsequently, in Table 6, we show the mapping from the likes ratio to the three categories, namely, low, medium, and high. We choose thresholds at 33% percentiles of the likes ratio distribution so that equal proportion of data points go in the three categories to maintain class balance. In the next step, we use Naive Bayes as our classification algorithm using features extracted from the text of YouTube video title as outlined in Table 3. Like before, we build three separate models using only sentiment, only named entity, and only duration & title length, and obtain accuracy of 37%, 47%, and 46%, respectively. However, the best accuracy of 92% is obtained (confusion matrix is depicted in Figure 7, observe that the model is able to precisely predict 96% of highly liked videos) when all the features are combined together, with video duration (*duration*) and named entity count (*tot_person* and *tot_org*) being most important features. Particularly, for most liked videos, video duration (*duration*) in the range of 2-3 minutes, is an important factor. So, we conclude that predicting likes ratios is hard, but predicting whether a YouTube video appearing in a fake news incident will receive a low, medium, or a high number of likes is possible with high accuracy.

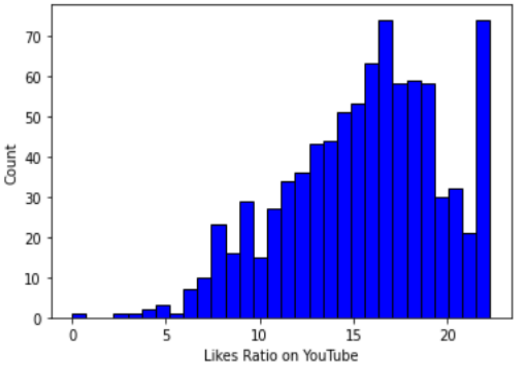


Figure 6: Distribution of Likes Ratio of videos on YouTube appearing in fake news incidents, and subsequently in Table 6, we categorize likes ratio into three categories.

Table 6: Table shows the Likes ratio range and the associated labels

Likes Ratio Range	**Category**
≤ 14	low
> 14 and < 17.6	medium
≥ 17.6	high

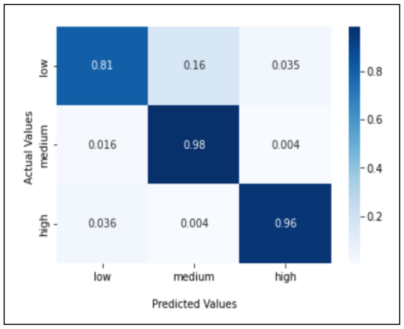


Figure 7: Confusion matrix for YouTube likes ratio model

To answer H2, we performed regression experiments to predict engagement rate ($ER_{YouTube}$) using the same features as described for like ratio above. As before, we did not get promising results. On performing K-S test on engagement rate distribution, we obtain test statistic of 0.550 with p-value less than 0.5 which indicates that it does not follow normal distribution. Therefore, we convert the engagement rate (distribution shown in Figure 8) also into three categories, namely, low, medium, and high (as depicted in Table 7) using 33% percentile as thresholds so that equal proportion of data goes into the three classes. Similar to what we did earlier, we ran the Naive Bayes algorithm using features extracted from the text of YouTube video title as outlined in Table 3 to predict these three classes and

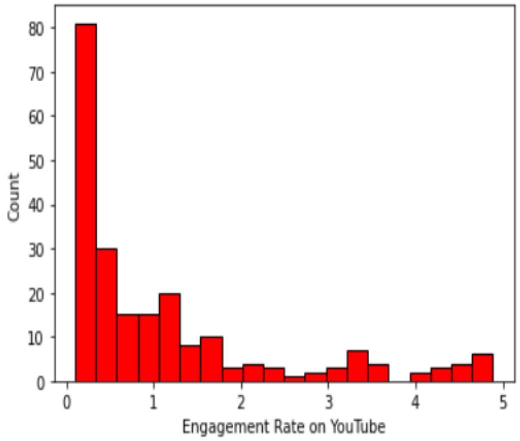


Figure 8: Distribution of Engagement Rate on YouTube videos appearing in fake news incidents, subsequently, in Table 7, we categorize videos into three categories.

Table 7: Engagement rate range and the associated categories

Engagement rate Range	**Category**
≤ 0.03	low
> 0.03 and < 3	medium
≥ 3	high

obtain 86% accuracy (Figure 9 depict the confusion matrix which shows that 93% of times highly engaging videos are correctly predicted by the model) when all features are used together. Features related to named entity count (tot_person and tot_org) turned out to be most important. Low accuracy of 36%, 47%, and 34% are obtained when only sentiment features, only named entity count based features, and duration & title length are used, respectively. So, we conclude that it is easier to predict the category of engagement rate for YouTube videos appearing in fake news incidents.

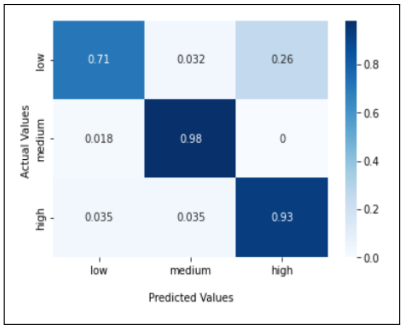
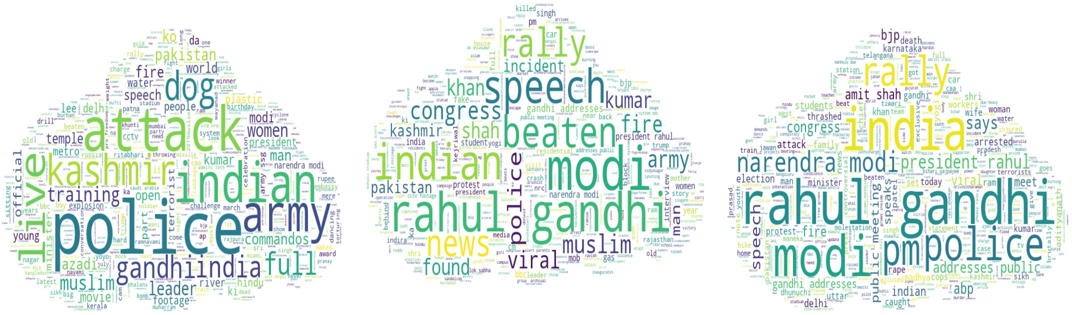


Figure 9: Confusion matrix for YouTube engagement rate model

Given that count of named entities play an important role in predicting likes ration and engagement count, in Figure 10, we depict the frequently used words found in YouTube video titles separately for videos that receive low, medium, and high engagement. We find that named entity words like police, attack, Kashmir are frequently found in highly engaging fake videos. We observe words like Congress, BJP, beaten, and fire in fake videos that received a medium level of engagement. In low-engaging fake videos, we find names of politicians like Rahul Gandhi and Narendra Modi. Although fake news incidents revolve around political themes, the presence of well-known politicians in fake videos makes them less engaging.



(a) High Engagement Videos (b) Medium Engagement Videos (c) Low Engagement Rate

Figure 10: Word Clouds of words found in YouTube video title.

5.2. Impact Assessment on Twitter

In the second part, we evaluate the performance of models that use features derived from tweet text and image, if present in the tweet, to predict likes-ratio ($LR_{Twitter}$) and engagement rate ($ER_{Twitter}$). Specifically, we ask two hypotheses as below.

- H3: Can text and image-based features derived from tweet predict **likes ratio** received on that tweet?
- H4: Can text and image-based features derived from tweet predict **engagement rate** received on that tweet?

To answer H3, we initially construct linear regression models to predict the likes ratio ($LR_{Twitter}$). However, the results did not turn out well as also observed earlier with likes ratio. Likes ratio does not follow normal distribution. K-S test for normality returns 0.786 as test statistic with p-value of 0 indicating that distribution of likes ratio is not normal distribution. So, following the approach employed in YouTube impact assessment, we converted the likes ratio, distribution depicted in Figure 11, into three categories (Table 8), namely, low, medium, and high using 33% percentile like ratio distribution.

For features, we use the sentiment scores (mixed_score, pos_score, neut_score, neg_score) and count of entities (e.g. locations, organizations, and more) present in each tweet text. Recall that for the tweets which have images, we use SightEngine API and Amazon rekognition API to extract the features (Table 4) from tweet images. Table 8 depicts the likes ratio range for those tweets appearing in fake news incidents that contain the image. We get an

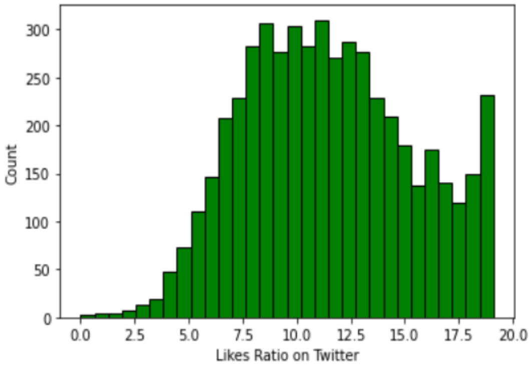


Figure 11: Distribution of Likes Ratio of tweets on Twitter appearing in fake news incidents, and subsequently in Table 8, we categorize likes ratio into three categories.

Table 8: Range of Likes Ratio ($LR_{Twitter}$) for tweets in Twitter and the associated categories. Features are derived from tweet text and image present in tweet.

Features Used	**Category**	**Likes Ratio Range**
Sentiment scores (mixed_score, pos_score, neut_score, neg_score) and count of entities (eg. locations, organizations, etc) in text	low medium high	≤ 9.42 > 9.42 & < 13.13 ≥ 13.13
Emotions, nudity, violence expressed in image in the tweet along with quality of image	low medium high	≤ 9.53 > 9.53 & < 13.21 ≥ 13.21

average accuracy of 37% in predicting categories (low, medium, and high) of likes ratio.

To answer H4, we converted the engagement rates (distribution shown in Figure 12 does not follow normal distribution as per K-S test, with p-value of 0 with test statistic of 0.816) into three categories (Table 9) using a 33% threshold in the distribution of engagement rates received on Twitter tweets appearing in fake news incidents. We use features extracted from sentiment and entities expressed in tweets appearing in fake news incidents in the first experiment. The naive Bayes algorithm gives 41% average accuracy in predicting engagement rate categories. In the second experiment, we leverage features related to emotions, violence, nudity, and image quality. In this case, we get an average of 37% using the Naive Bayes algorithm. We understand

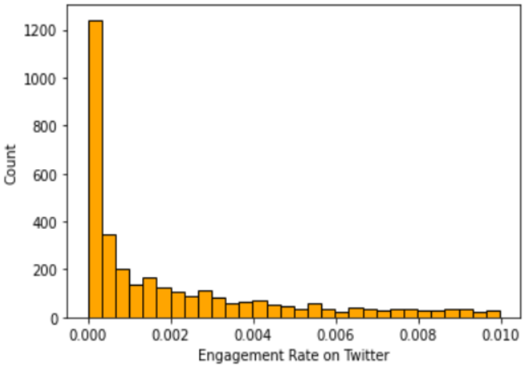


Figure 12: Distribution of Engagement Rate of tweets on Twitter appearing in fake news incidents, and subsequently in Table 9, we categorize likes ratio into three categories.

Table 9: Engagement Rate $ER_{Twitter}$ range of tweets in Twitter and the associated categories. Features are derived from tweet text and image present in tweet.

Features Used	**Category**	**Engagement Rate Range**
Sentiment scores (mixed_score, pos_score, neut_score, neg_score) and count of entities (eg. locations, organizations, etc) in text	low	$\leq$ 7.8e-4
	medium	$>$ 7.8e-4 & $<$ 9.7e-3
	high	$\geq$ 9.7e-3
Emotions, nudity, violence expressed in image in the tweet along with quality of image	low	$\leq$ 1.37e-3
	medium	$>$ 1.37e-3 & $<$ 1.45e-2
	high	$\geq$ 1.45e-2

that this baseline accuracy is a near-random guess and far from acceptable. This means that features derived from fake news incident articles are not good predictors for the impact of tweets on Twitter. Furthermore, in the future, features should be derived from the tweet itself using Twitter API, provided the fake tweet has not been removed.

6. Open Research Issues

Based on our experience of working on fake news, we suggest the following open research issues that require attention.

- All fake incidents reported by fact-checkers are not fake stories. Some of them are also just narratives, opinions, and views. For instance, this[13] story reported by OpIndia is an opinion expressed by the fact-checker that an ex-diplomat in India has made a tweet in a manner that belittles the Prime Minister of India.

- Some of the fact-checkers could suffer from coverage and selection biases. Owing to various challenges, it appears that they do not cover all fake news incident stories and their coverage could be biased. More scientific investigation needs to be done in this direction. If true, then this would be a disturbing trend because it jeopardizes the fight against fake news that needs to be fair and credible.

- The same fake news story can be reported by two or more fact-checkers, so a computational approach is needed to identify and remove (or merge) these duplicates. Identifying duplicates is a non-trivial task because each fact-checker would report fake stories in their editorial style. In this work, we adopt a naive approach of removing the stop words, and then comparing common words in the titles of fake news incidents, however, a more robust approach is desirable.

- In the fact news incident (article), the text of the article is written in free-form English language, and it contains URLs pointing to tweets, videos, Facebook posts, WhatsApp screenshots, and many more. However, only some of these URLs point to the fake posts; other URLs could be refuting the fake posts or comprise evidence and explanations given by fact-checkers to refute the fake news. So, a computational pipeline for automated retrieval of correct URLs that point to fake posts needs to be developed. Further, a generic taxonomy needs to be identified (and defined) based on fact news incident articles. For instance, it identifies the victim of fake news, initiator (first post) of fake news in OSM, spreader of fake posts, the content of the fake post, and more.

- Many URLs in fake news incidents that point to the fake posts on OSM platforms get either removed by the user or by the platform. So, analyzing and measuring their impact in OSM platforms becomes challenging. To study them, we would require an in-depth analysis of

[13] https://www.opindia.com/2019/10/former-diplomat-kc-singh-spreads-fake-news-sandeep-dhaliwal-houston-nrg-stadium-belittle-howdy-modi/

deleted fake tweets using the archived Twitter stream. [14]

- Fake news incidents (stories) reported by fact-checkers on their web page also have an option to share this story on OSM platforms. It will be interesting to perform a characterization study on who usually shares these stories, how users react to it on the OSM platform, accept the fact-checker version, or refuse to believe it is fake news.
- There is a digital divide in societies. Some technically savvy people can create fake news to evade detection. And then some people are not too tech-savvy and less literate, who usually fall for the fake post and like/share it. So, a controlled user experiment can be done to assess the susceptibility of users to fall for fake posts. It will help in creating awareness in society to fight fake news.
- It would be interesting to investigate and compare the impact of fake news incidents with real news incidents on social media platforms. It will also help in building models that generalize in both real and fake news incidents.

7. Conclusion & Future Scope of our work

The spread of fake news raises concerns all over the world. Fake news is causing unrest in the society, loss of life and property. As a result, many fact-checking websites (fact-checkers) have emerged to identify and detect fake news incidents on social media platforms. We have created a novel dataset, **FakeNewsIndia**, comprising 4,803 fake news incidents in the Indian context in our work. We plan to use the proposed pipeline to keep updating the fake news incidents reported by fact-checkers. So far, we have 5,031 tweets on Twitter, and 866 videos on YouTube mentioned in 4,803 fake news incidents. Subsequently, we explored the question *whether text and image-based features appearing in tweets and video titles can help us in predicting the impact of fake news incidents on social media platforms.* To this end, we find that we can predict the impact (traction) of video on YouTube with 92% and 86% accuracy, as measured using likes ratio and engagement rate categories, respectively. However, the same cannot be said for tweets on Twitter. We

[14] https://archive.org/details/twitterstream

could only achieve accuracy of 37% and 41% in predicting the impact of fake news incidents on Twitter using likes ratio and engagement rate categories as our impact metrics, respectively. Low impact prediction means that it is challenging to measure the popularity (impact) of a tweet related to a fake news incident, and therefore, it is difficult to initiate an early response. Based on this work, we suggest the following directions in the future.

- We shall use the proposed collection approach to continue to collect more fake news incidents periodically. More the data better would be the predictive capability of the models. The first version of the dataset is available at this[15] link.
- We did not focus on fake news incidents in regional languages. However, while working on this problem, we found Singhal et al. [24] have collected numerous fake news incidents in regional languages in India.
- At an algorithmic level, we can leverage a more advanced machine learning algorithm beyond naive bayes that we primarily use for impact assessment. Further, feature extraction image-based features using convolutional neural network-based architecture can get richer features.
- We can broaden the scope of our work and compare fake news incidents with real news incidents. It would be worth exploring the propagation patterns and impact (popularity) of fake and real news incidents.

References

[1] A. Mayfield, What is social media (2008).

[2] M. Kang, Measuring social media credibility: A study on a measure of blog credibility, Institute for Public Relations (2010) 59–68.

[3] D. Westerman, P. R. Spence, B. Van Der Heide, Social media as information source: Recency of updates and credibility of information, Journal of computer-mediated communication 19 (2) (2014) 171–183.

[4] L. Graves, Deciding what's true: Fact-checking journalism and the new ecology of news, Ph.D. thesis, Columbia University (2013).

[15] https://github.com/rishabhkaushal/fakenewsincidentsIndia

[5] L. Graves, M. A. Amazeen, Fact-checking as idea and practice in journalism, in: Oxford Research Encyclopedia of Communication, 2019.

[6] T. Gillespie, Content moderation, ai, and the question of scale, Big Data & Society 7 (2) (2020) 2053951720943234.

[7] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, et al., The science of fake news, Science 359 (6380) (2018) 1094–1096.

[8] X. Zhou, R. Zafarani, K. Shu, H. Liu, Fake news: Fundamental theories, detection strategies and challenges, in: Proceedings of the twelfth ACM international conference on web search and data mining, 2019, pp. 836–837.

[9] M. Aldwairi, A. Alwahedi, Detecting fake news in social media networks, Procedia Computer Science 141 (2018) 215–222.

[10] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media, arXiv preprint arXiv:1809.01286 (2018).

[11] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, Journal of economic perspectives 31 (2) (2017) 211–36.

[12] R. N. Spicer, Lies, damn lies, alternative facts, fake news, propaganda, pinocchios, pants on fire, disinformation, misinformation, post-truth, data, and statistics, in: Free Speech and False Speech, Springer, 2018, pp. 1–31.

[13] R. R. Mourão, C. T. Robertson, Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information, Journalism Studies 20 (14) (2019) 2077–2095.

[14] A. J. Mills, C. Pitt, S. L. Ferguson, The relationship between fake news and advertising: brand management in the era of programmatic advertising and prolific falsehood, Journal of Advertising Research 59 (1) (2019) 3–8.

[15] D. N. Rapp, N. A. Salovich, Can't we just disregard fake news? the consequences of exposure to inaccurate information, Policy Insights from the Behavioral and Brain Sciences 5 (2) (2018) 232–239.

[16] T. Mitra, E. Gilbert, Credbank: A large-scale social media corpus with associated credibility annotations, in: Ninth International AAAI Conference on Web and Social Media, 2015, pp. 121–139.

[17] W. Y. Wang, " liar, liar pants on fire": A new benchmark dataset for fake news detection, arXiv preprint arXiv:1705.00648 (2017).

[18] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyperpartisan and fake news, arXiv preprint arXiv:1702.05638 (2017).

[19] S. Kemp, Digital 2021: India.
URL `https://datareportal.com/reports/digital-2021-india`

[20] D. Welle, India fake news problem fueled by digital illiteracy.
URL `https://www.taiwannews.com.tw/en/news/4140436`

[21] K. K. Aldous, J. An, B. J. Jansen, Predicting audience engagement across social media platforms in the news domain, in: International Conference on Social Informatics, Springer, 2019, pp. 173–187.

[22] G. Santia, J. Williams, Buzzface: A news veracity dataset with facebook user commentary and egos, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 12, 2018.

[23] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, L. de Alfaro, Some like it hoax: Automated fake news detection in social networks, arXiv preprint arXiv:1704.07506 (2017).

[24] S. Singhal, R. R. Shah, P. Kumaraguru, Factorization of fact-checks for low resource indian languages, arXiv preprint arXiv:2102.11276 (2021).

[25] L. Wu, F. Morstatter, K. M. Carley, H. Liu, Misinformation in social media: definition, manipulation, and detection, ACM SIGKDD Explorations Newsletter 21 (2) (2019) 80–90.

[26] F. Yu, Q. Liu, S. Wu, L. Wang, T. Tan, et al., A convolutional approach for misinformation identification., in: IJCAI, 2017, pp. 3901–3907.

[27] S. Jain, V. Sharma, R. Kaushal, Towards automated real-time detection of misinformation on twitter, in: 2016 International conference on advances in computing, communications and informatics (ICACCI), IEEE, 2016, pp. 2015–2020.

[28] P. Assiroj, A. N. Hidayanto, H. Prabowo, H. L. H. S. Warnars, et al., Hoax news detection on social media: A survey, in: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), IEEE, 2018, pp. 186–191.

[29] F. Tchakounté, K. A. Calvin, A. A. A. Ari, D. J. F. Mbogne, A smart contract logic to reduce hoax propagation across social media, Journal of King Saud University-Computer and Information Sciences (2020).

[30] A. Zubiaga, A. Jiang, Early detection of social media hoaxes at scale, ACM Transactions on the Web (TWEB) 14 (4) (2020) 1–23.

[31] V. Qazvinian, E. Rosengren, D. Radev, Q. Mei, Rumor has it: Identifying misinformation in microblogs, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011, pp. 1589–1599.

[32] T. Takahashi, N. Igata, Rumor detection on twitter, in: The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems, IEEE, 2012, pp. 452–457.

[33] R. Dayani, N. Chhabra, T. Kadian, R. Kaushal, Rumor detection in twitter: An analysis in retrospect, in: 2015 IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS), IEEE, 2015, pp. 1–3.

[34] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD Explorations Newsletter 19 (1) (2017) 22–36.

[35] F. Monti, F. Frasca, D. Eynard, D. Mannion, M. M. Bronstein, Fake news detection on social media using geometric deep learning, arXiv preprint arXiv:1902.06673 (2019).

[36] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, H. Liu, Unsupervised fake news detection on social media: A generative approach, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 33, 2019, pp. 5644–5651.

[37] E. Okoro, B. Abara, A. Umagba, A. Ajonye, Z. Isa, A hybrid approach to fake news detection on social media, Nigerian Journal of Technology 37 (2) (2018) 454–462.

[38] C. Guo, J. Cao, X. Zhang, K. Shu, M. Yu, Exploiting emotions for fake news detection on social media, arXiv preprint arXiv:1903.01728 (2019).

[39] C. Guo, J. Cao, X. Zhang, K. Shu, H. Liu, Dean: Learning dual emotion for fake news detection on social media, arXiv preprint arXiv:1903.01728 (2019).

[40] K. Stahl, Fake news detection in social media, California State University Stanislaus 6 (2018) 4–15.

[41] S. Tschiatschek, A. Singla, M. Gomez Rodriguez, A. Merchant, A. Krause, Fake news detection in social networks via crowd signals, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 517–524.

[42] Y.-J. Lu, C.-T. Li, Gcan: Graph-aware co-attention networks for explainable fake news detection on social media, arXiv preprint arXiv:2004.11648 (2020).

[43] K. Shu, S. Wang, H. Liu, Understanding user profiles on social media for fake news detection, in: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE, 2018, pp. 430–435.

[44] Y. Liu, Y.-F. B. Wu, Fned: a deep network for fake news early detection on social media, ACM Transactions on Information Systems (TOIS) 38 (3) (2020) 1–33.

[45] Q. Zhang, S. Zhang, J. Dong, J. Xiong, X. Cheng, Automatic detection of rumor on social network, in: Natural Language Processing and Chinese Computing, Springer, 2015, pp. 113–122.

[46] S. Kwon, M. Cha, K. Jung, Rumor detection over varying time windows, PloS one 12 (1) (2017) e0168344.

[47] J. Ma, W. Gao, K.-F. Wong, Rumor detection on twitter with tree-structured recursive neural networks, Association for Computational Linguistics, 2018.

[48] K. Swani, G. R. Milne, B. P. Brown, A. G. Assaf, N. Donthu, What messages to post? evaluating the popularity of social media communications in business versus consumer markets, Industrial Marketing Management 62 (2017) 77–87.

[49] S. Mishra, M.-A. Rizoiu, L. Xie, Modeling popularity in asynchronous social media streams with recurrent neural networks, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 12, 2018.

[50] J. Xu, M. Van Der Schaar, J. Liu, H. Li, Forecasting popularity of videos using social media, IEEE Journal of Selected Topics in Signal Processing 9 (2) (2014) 330–343.

[51] K. Shu, D. Mahudeswaran, H. Liu, Fakenewstracker: a tool for fake news collection, detection, and visualization, Computational and Mathematical Organization Theory 25 (1) (2019) 60–71.