Social Computing Research in India

Feb 18, 2023
Research Career in Computer Science
IIT Palakkad

Ponnurangam Kumaraguru (“PK”)
#ProfGiri CS IIIT Hyderabad
ACM Distinguished Member
TEDx Speaker

https://www.instagram.com/pk.profgiri/

https://www.linkedin.com/in/ponguru/

Ponnurangam Kumaraguru “PK”
@ponguru

What is Social Computing?

Social computing is an area of computer science that is concerned with the intersection of social behavior and computational systems. It is based on creating or recreating social conventions and social contexts through the use of software and technology. Thus, blogs, email, instant messaging, social network services, wikis, social bookmarking and other instances of what is often called social software illustrate ideas from social computing.

https://en.wikipedia.org/wiki/Social_computing

Our overwhelmed legal system

The long pendency of cases raises questions about the accountability of the legal edifice to citizens

At the end of 2018, there was a shortage of more than 5,000 judges in India's lower courts, usually the first point of contact between the people and the judiciary
(Shutterstock)

Tech may help remove barriers in using local languages in court: CJI NV Ramana

The CJI was speaking after laying the foundation stone for a nine-storey administrative block of Madras HC, and inaugurating combined court buildings in Kallakurichi and Namakkal.

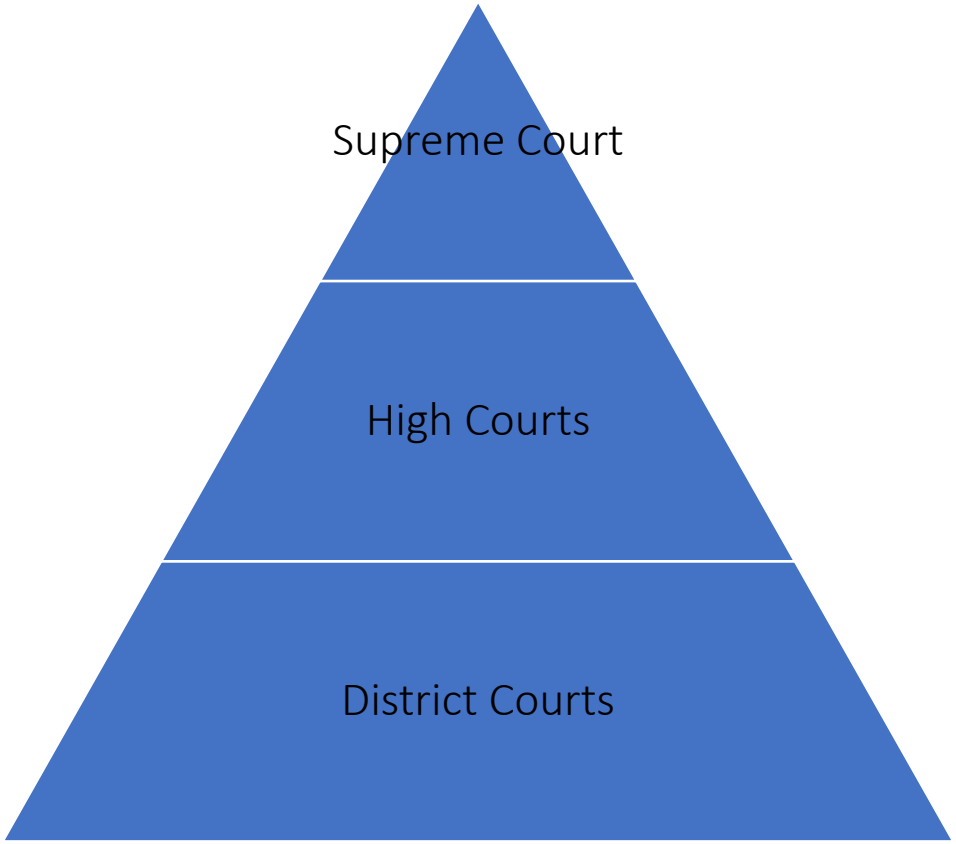
Published: 24th April 2022 06:06 AM | Last Updated: 24th April 2022 06:06 AM | A+ A A-

Legal AI for Indian Context

District courts are usually the first point of contact between the people and the judiciary.

Lower courts in India are burdened with a backlog of cases (~40 million as of 2021).

Local languages used in the documents filed in district courts in India.

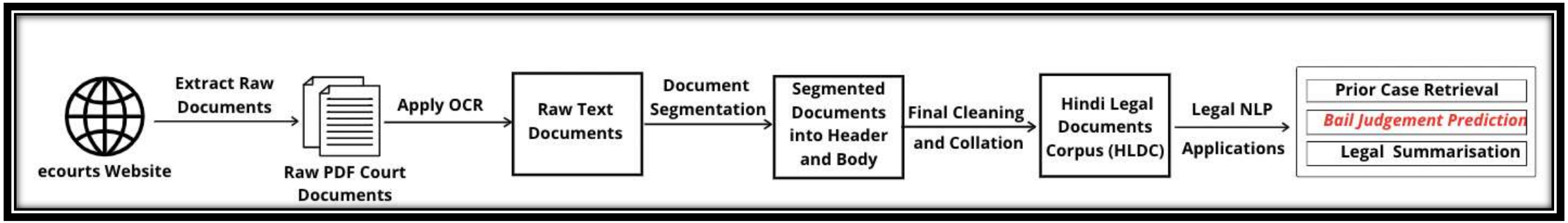


Legal AI / NLP - Data

We collected ~900k district court case documents from Uttar Pradesh

All documents in Hindi, written in Devanagari

There are legal corpora for European Court of Justice and Chinese courts, none for Indian district courts



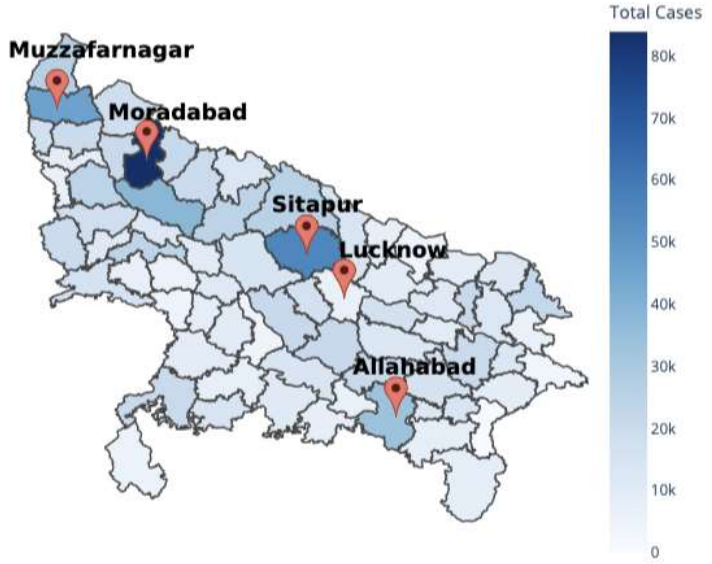
Legal AI / NLP - Data

There are around 300 different case types, table shows the prominent ones

Majority of the case documents correspond to Bail Applications

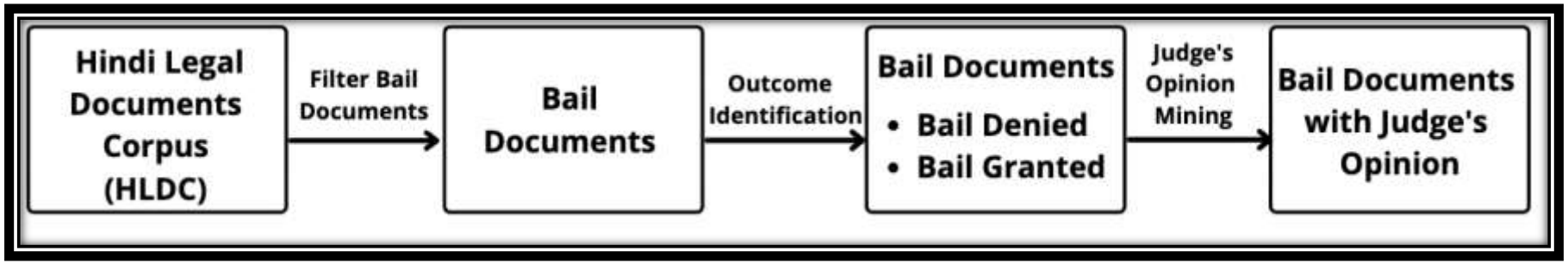
Case Type	% of Cases in HLDC
Bail Applications	31.71
Criminal Cases	20.41
Original Suits	6.54
Warrant or Summons in Criminal Cases	5.24
Complaint Cases	4.37
Civil Misc	3.4
Final Report	3.32
Civil Cases	3.23
Others (Matrimonial Cases, Session Trial, Motor Vehicle Act, etc.)	21.75

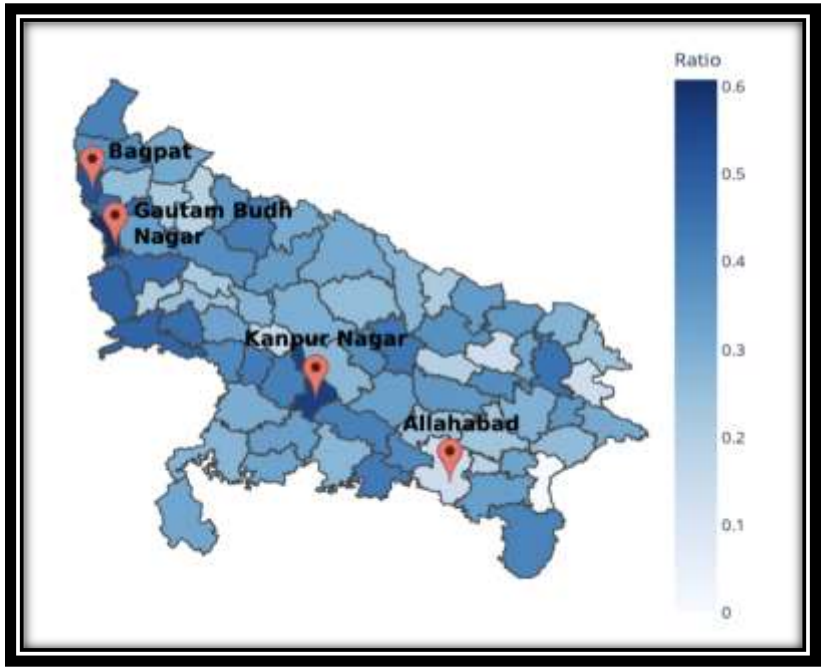
Case types in HLDC



Variation in number of case documents per district

Legal AI / NLP - Bail Documents



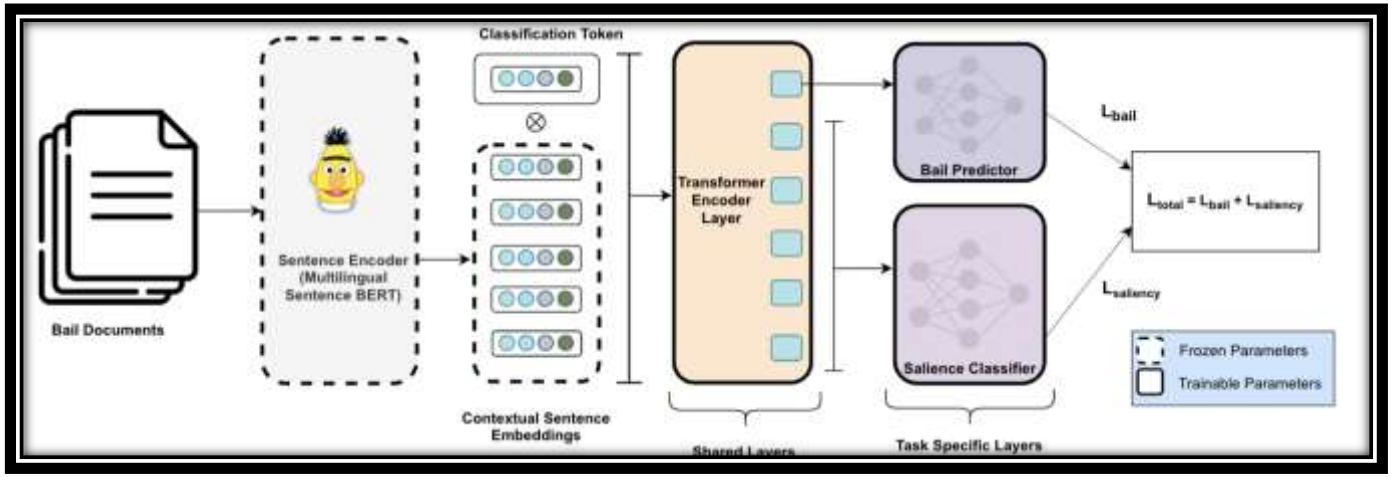


District-wise ratio of number of bail applications to total cases

Legal AI / NLP - Bail Prediction Model

In general, the performance is lower in district-wise settings, possibly due to large variation across districts

Overall, summarization models perform better than Doc2Vec and simpler Transformer-based models



Model Name	District-wise		All Districts	
	Acc.	F1	Acc.	F1
Doc2Vec + SVM	0.72	0.69	0.79	0.77
Doc2Vec + XGBoost	0.68	0.59	0.67	0.57
IndicBert-First 512	0.65	0.62	0.73	0.71
IndicBert-Last 512	0.62	0.60	0.78	0.76
TF-IDF+IndicBert	0.76	0.74	**0.82**	**0.81**
TextRank+IndicBert	0.76	0.74	0.82	0.81
Salience Pred.+IndicBert	0.76	0.74	0.80	0.78
Multi-Task	**0.78**	**0.77**	*0.80*	*0.78*

Legal AI / NLP for Indian Context

HLDC: Hindi Legal Documents Corpus

Arnav Kapoor[†], Mudit Dhawan[‡], Anmol Goel[†], T.H. Arjun[†], Akshala Bhatnagar[‡]
Vibhu Agrawal[‡], Amul Agrawal[†], Arnab Bhattacharya[¶]
Ponnurangam Kumaraguru[†], Ashutosh Modi[¶]
[†]IIIT Hyderabad,[‡]IIIT Delhi,[¶]IIT Kanpur
{arnav.kapoor, anmol.goel, arjun.thekoot}@research.iiit.ac.in
{mudit18159, akshala18012, vibhu18116}@iiitd.ac.in
amul.agrawal@students.iiit.ac.in, pk.guru@iiit.ac.in
{arnabb,ashutoshm}@cse.iitk.ac.in

HLDC: Hindi Legal Documents Corpus

Legal AI / NLP for Indian Context - Takeaways

Indian Legal documents are a rich a source of domain-specific Indic-language corpora, readily available online

Multiple tasks still need attention especially for Indian settings

- Legal Summarization
- Case recommendations
- Citation predictions / network

Sleeping beauty

Bias

Where to start if you are interested?

Saptarshi Ghosh https://sites.google.com/site/saptarshighosh/

Ashutosh Modi https://ashutosh-modi.github.io/

Prathamesh Kalamkar https://scholar.google.com/citations?user=7RaVib0AAAAJ&hl=en&oi=ao

OpenNyAI https://opennyai.org/

Ilias Chalkidis https://iliaschalkidis.github.io/

Off course, Precog ☺ https://precog.iiit.ac.in/

Vishva thakkar
@Vishva46911509

Follow

All set
Long day but very important day for all of us
लोकसभा चुनावों के लिए पहले चरण का मतदान शुरू
#VotingRound1
#LokSabhaElections2019

3:36 PM - 10 Apr 2019

Chintada Karthik
@ChintadaKarthi1

Follow

OKKADEY VACCHADU, 🚶 media support ledu,pedda pedda leaders leru,kaani nebadddadu
AP VOTES FOR GLASS

1.4K 4.8K 23K

Ravi Kumar singh
@RaviKum24798035

Follow

Replying to @kanhaiyakumar

Waise tumlogo to kahte ho Supreme Court election comition sb modi k isare p Kam krte hai.. Fir a nhi kahna evm hack ho gya

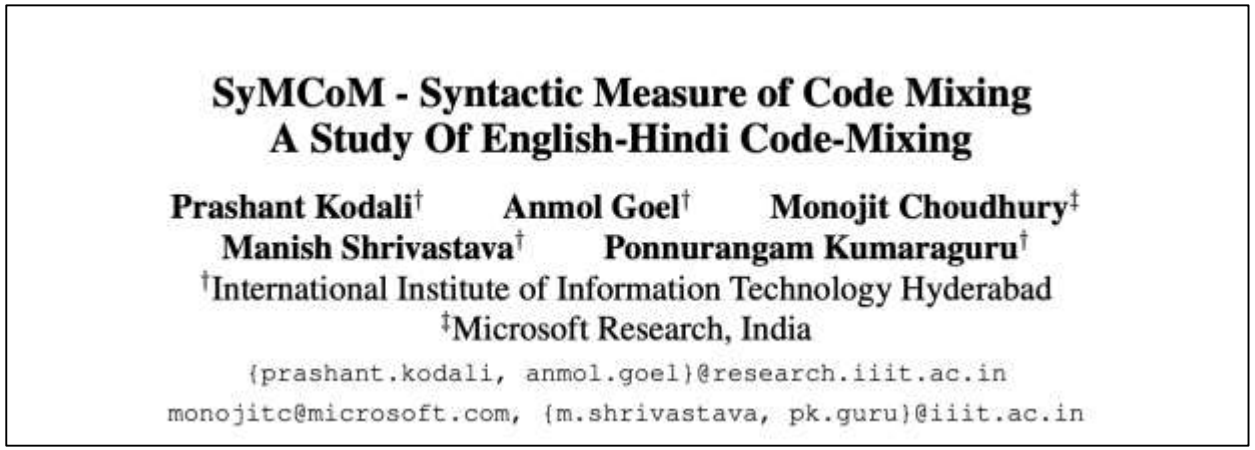
3:49 PM - 10 Apr 2019

Code-mix computationally challenging hai.

Predominantly noticed in social networks and speech data

Social Media text processing poses certain challenges

- @, #, https://
- Incorrect Spelling & Romanisation
- Mixing two + languages – Hinglish



SyMCoM - Syntactic Measure of Code Mixing
A Study Of English-Hindi Code-Mixing

Prashant Kodali[†] **Anmol Goel**[†] **Monojit Choudhury**[‡]
Manish Shrivastava[†] **Ponnurangam Kumaraguru**[†]
[†]International Institute of Information Technology Hyderabad
[‡]Microsoft Research, India
{prashant.kodali, anmol.goel}@research.iiit.ac.in
monojitc@microsoft.com, {m.shrivastava, pk.guru}@iiit.ac.in

[Paper at ACL Findings 2022]

Problems / Applications

Identify code mix sentences in wild

Generating natural code mix sentences

Downstream application in dialog systems, human computer interface

Making sense of the social content can help
making choices, recommendations, civic services

Where to start if you are interested?

Monojit Choudhury https://www.microsoft.com/en-us/research/people/monojitc/

Preethi Joyti https://www.cse.iitb.ac.in/~pjyothi/

Vivek Srivastava https://sites.google.com/view/vivek-srivastava/

Pushpak Bhattacharyya https://www.cse.iitb.ac.in/~pb/

Thamar Solorio http://solorio.uh.edu/

Off course, Precog ☺ https://precog.iiit.ac.in/


Chowkidar Amit Shah
@AmitShah
National President of Bharatiya Janata Party / President, Gujarat Cricket Association


Chowkidar Arun Jaitley
@arunjaitley
Minister of Finance and Corporate Affairs, Government of India

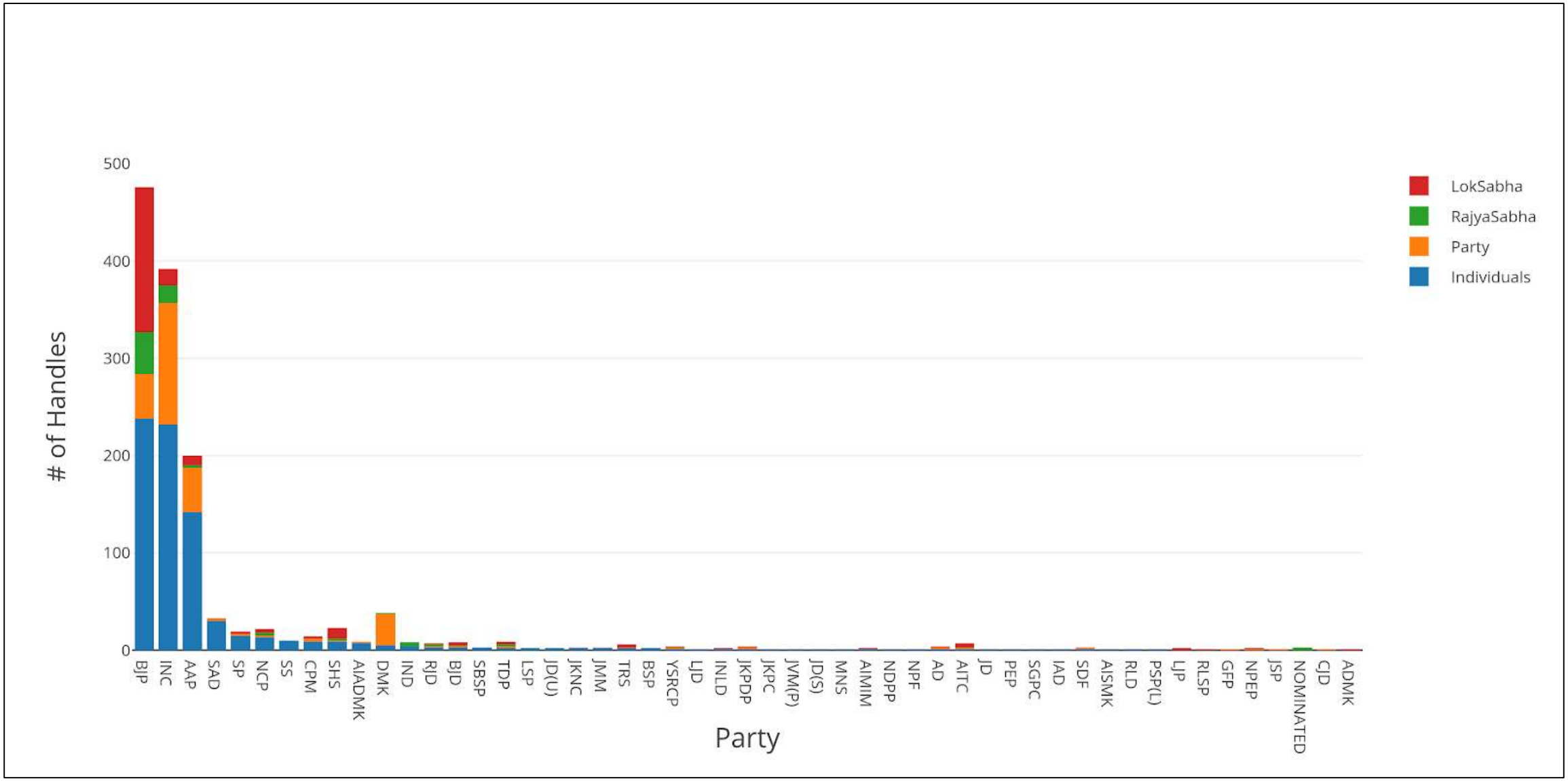

Chowkidar Smriti Z Irani
@smritiirani
Union Cabinet Minister of Textiles, Government of India. Member of Parliament, Rajya Sabha representing the state of Gujarat.


Chowkidar Jayesh Mehta
@JMehta65
Follower of Prime Minister of India, Narendra Modi. #ModiFor2019


Chowkidar ramakrishna nannuta
@rknannuta
Soldier of #NaMo, Proud Indian, Prime Minister of India Shri @narendramodi follows me,Non Veg Addict, Master Chef,Senior Technical Lead,Openstack Cloud


Chowkidar Haritha
@csharitha
Proud Indian. Blessed to be followed by Hon'ble PM Narendra Modi ji & @ramkripalmp & @byadavbjp. Company Secretary. Indian Army Fan. Jai Hind!

Verified handles: 71 (2014), 1,268 (2019)



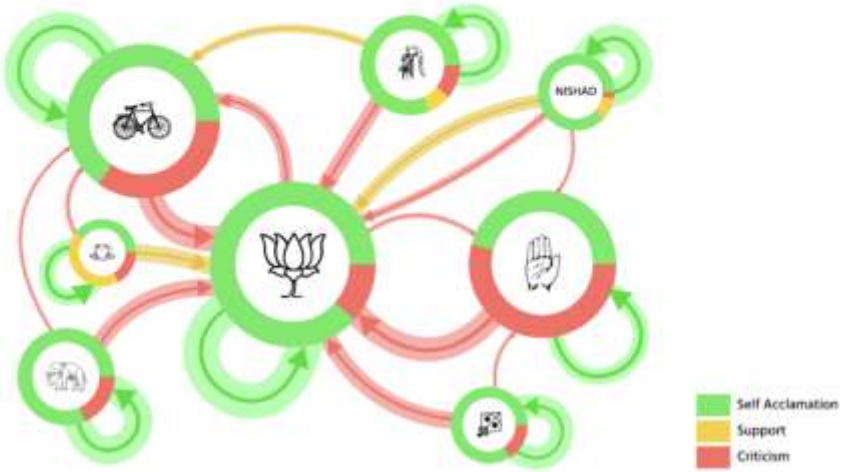
UP ELECTIONS 2022

What #hashtag wars tell us about the organization of Uttar Pradesh IT cells, and how parties mark their allies and enemies

February 9, 2022

Pitched battles on Twitter are increasingly common among the key players in the Uttar Pradesh elections, since the praises, innuendo, and insults that take shape and virality on Twitter often find new lives on other social media such as WhatsApp, but also on mainstream television and web-based news. We studied nine parties contesting the state legislative elections in Uttar Pradesh – BJP, SP, INC, RLD, BSP, AIMIM, Nishad Party, AAP, and Apna Dal, and collected the hashtags used by over 11,000 politicians at various levels, from sitting MPs and MLAs to party workers. We collected all the hashtags used by the said politicians during Sep 2021 to Jan 2022.

We sorted them based on the frequency of their usage, and listed the top 10 hashtags for each week, hand-coded each of them for the textual content.We see three forms of critical hashtags — first, a character assassination, typically aimed at an individual politician (such as namecalling the Gandhi family by referring to Jawaharlal Nehru as a womanizer through hashtags like #डरकी_दियस,), second a criticism of a party's ideology or casting of an entire party under a specific form of name-calling (such as questioning the patriotism of Samajwadi party through hashtags like #जिन्ना_प्रेमी_अखिलेश), and third, an attack on some action attributable to the party (such as blaming unemployment or the farming crisis on on BJP's policies through hashtags like #किसान_हत्यारी_भाजपा). We also find two broad directions of attribution in the hashtags. The first is a hashtag construction that attempts to show one's own leader or party in a positive light – which can we refer to as self-acclaim, or claims that one's party is going to win, and the second is antagonistic or critical hashtags that attempt to show one's opposition negatively. We visualize a third category here, of support, which suggests hashtags that are supportive of a third party, usually an alliance party.



UP ELECTIONS 2022

Uttar Pradesh Elections and Twitter: The Cast of Characters

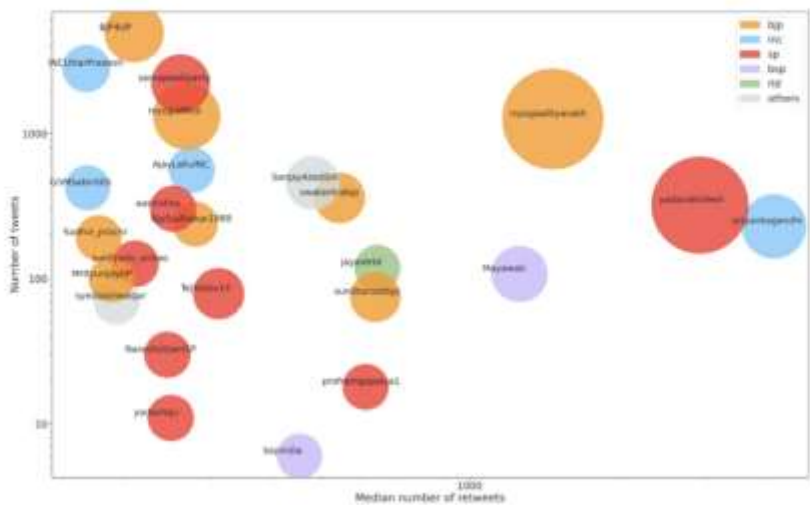
January 27, 2022

The Uttar Pradesh elections are being fought on social media just as much as they are in the physical spaces of the campaign. While messages forwarded on WhatsApp are a significant piece of what drives peoples' political opinions, with every major party investing in building WhatsApp groups of supporters and likely voters, and flooding them with outreach material, the primary means of official branding continues to be Twitter for most of the parties and major leaders within the parties.

Over the coming weeks, a collaboration between the University of Michigan and IIIT-Hyderabad will analyse the use of social media by politicians on various aspects of the electoral campaign in Uttar Pradesh. As a starting point, we are releasing the methodology and an introduction to the key players and parties. In addition, we will provide an interactive dashboard for users to visualize trends and analyze the performance of various parties and candidates. This is the first in a series of blogposts on the interesting narratives shaping up online. We will also be making the data public in coming weeks.

Key players

We visualised the key players across parties based on their following and appeal in the state. First, we look at a visualisation of those politicians who are actively part of the election either in their party roles (such as general secretary or in-charge for elections) or as MLA candidates or hopefuls. We excluded sitting MPs, except those actively part of the process for ease of visualisation.



UP ELECTIONS 2022

Midpoint Melodrama: UP Political Twitter Halfway through Polling Phases

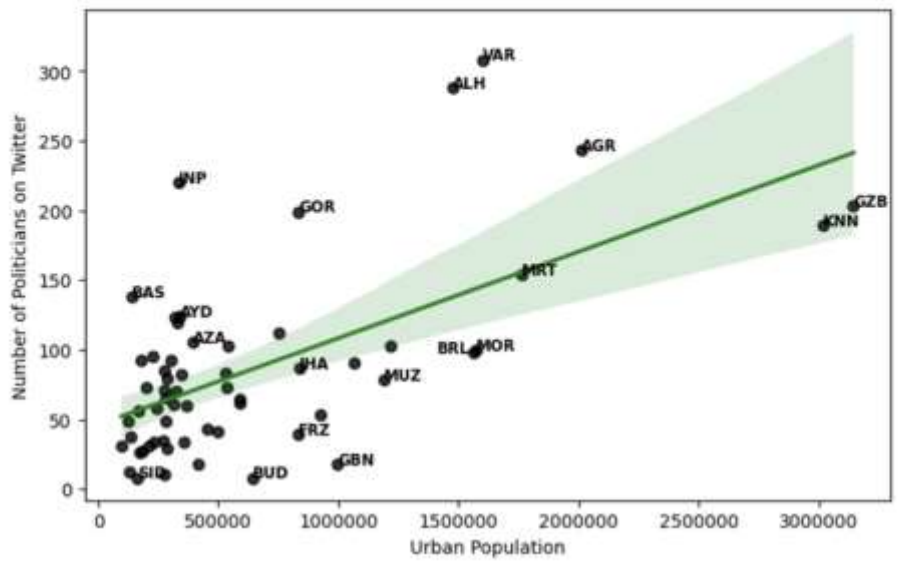
March 7, 2022

We examined the use of Twitter by 10426 politicians from Uttar Pradesh, across all levels of seniority, midway into the phases of polling in the Uttar Pradesh elections of 2022, starting January 1, 2022, and up until March 1, 2022. We summarize here our key findings by examining four key questions.

How does Twitter relate to Urbanization and Literacy?

To address the first question, we examined the number of urban residents by location, and levels of literacy, and find there is a statistically significant relationship between both the number (and percentage) or urban residents and the likelihood that politicians in that area use Twitter, as well as with the level of literacy and the likelihood of tweeting. We removed Lucknow district from the equations we ran, since it is an outlier being the capital where most parties have a number of listed post holders.

We see some variation — for instance, the district with the highest urban populations – Ghaziabad, adjoining Delhi, and Kanpur, both have fewer politicians online than Varanasi, the Prime Minister's constituency, which has had a lot of mid- to lower-level politicians go online, and Allahabad / Prayagraj, which is a politically important district in the state, and home to several key politicians from across parties.



https://precog.iiit.ac.in/blog/

Is change the only constant? Profile change perspective on #LokSabhaElections2019

Kumari Neha
IIIT Delhi
nehak@iiitd.ac.in

Shashank Srikanth
IIIT Hyderabad
shashank.s@research.iiit.ac.in

Sonali Singhal
IIIT Delhi
sonali18317@iiitd.ac.in

Shwetanshu Singh
IIIT Delhi
shwetanshus@iiitd.ac.in

Arun Balaji Buduru
IIIT Delhi
arunb@iiitd.ac.in

Ponnurangam Kumaraguru
IIIT Delhi
pk@iiitd.ac.in

ABSTRACT

Users on Twitter are identified with the help of their profile attributes that consists of a username, display name, profile image, to name a few. The profile attributes that users adopt can reflect their interests, belief, or thematic inclinations. Literature has proposed the implications and significance of profile attribute change for a random population of users. However, the use of profile attribute for

In this work, we conduct a large-scale study of the profile change behavior of users on Twitter during the 2019 general elections in India. These elections are of interest from the perspective of social media analysis for many reasons. Firstly, the general elections in India were held in 7 phases from 11th April, 2019 to 19th May, 2019. Thus, the elections serve as a rich source of data for social movements and changing political alignments throughout the two

Hashtags are (not) judgemental: The untold story of Lok Sabha elections 2019

Saurabh Gupta
IIIT-Delhi
saurabhg@iiitd.ac.in

Asmit Kumar Singh
IIIT-Delhi
asmit18025@iiitd.ac.in

Arun Balaji Buduru
IIIT-Delhi
arunb@iiitd.ac.in

Ponnurangam Kumaraguru
IIIT-Delhi
pk@iiitd.ac.in

ABSTRACT

Hashtags in online social media have become a way for users to build communities around topics, promote opinions, and categorize messages. In the political context, hashtags on Twitter are used by users to campaign for their parties, spread news, or to get followers and get a general idea by following a discussion built around a hashtag. In the past, researchers have studied certain types and specific properties of hashtags by utilizing a lot of data collected around hashtags. In this paper, we perform a large-scale empirical analysis of elections using only the hashtags shared on Twitter during the 2019 Lok Sabha elections in India. We study the trends and events unfolded on the ground, the latent topics to uncover representative hashtags and semantic similarity to relate hashtags with the election outcomes. We collect over 24 million hashtags to perform extensive experiments. First, we find the trending hashtags to cross-reference them with the tweets in our dataset to list down notable events. Second, we use Latent Dirichlet Allocation to find topic patterns in the dataset. In the end, we use skip-gram word embedding model to find semantically similar hashtags. We propose popularity and an influence metric to predict election outcomes

1 INTRODUCTION

Online social media platforms like Twitter are being used by people to spread information and opinions among other users. A lot of times, people are observed reporting the ground events happening near them, making Twitter a source of getting breaking news [2, 26]. For example, when the terrorist attacks in Mumbai in 2008 were happening, Twitter users in India (especially in Mumbai) were providing an instant eyewitness account of what was happening at the ground [3]. More recently, a lot of media channels covered the reactions of people all over India using Twitter when article 370 was scraped [8, 9]. Twitter is considered so effective that even the Indian government recently asked them to remove accounts spreading rumors about Kashmir [24].

Hashtags on Twitter are generally used: by many brands to promote a product; by users to categorize their messages, build communities around a topic; and are an efficient way to join public discussions. It is very common to see the use of hashtags to aid the formation of ad-hoc publics around specific events like #ausvotes, #londonriots, #wikileaks[5], and #elections2019 (see Figure 1).

In the past, hashtags helped users connect to various political

Catching up with trends: The changing landscape of political discussions on twitter in 2014 and 2019

AVINASH TULASI *, IIIT Delhi
KANAY GUPTA *, IIIT Hyderabad
OMKAR GURJAR, IIIT Hyderabad
SATHVIK SANJEEV BUGGANA, IIIT Hyderabad
PARAS MEHAN, IIIT Delhi
ARUN BALAJI BUDURU, IIIT Delhi
PONNURANGAM KUMARAGURU, IIIT Delhi

The advent of 4G increased the usage of internet in India, which took a huge number of discussions online. Online Social Networks (OSNs) are the center of these discussions. During elections, political discussions constitute a significant portion of the trending topics on these networks. Politicians and political parties catch up with these trends, and social media then becomes a part of their publicity agenda. We cannot ignore this trend in any election, be it the U.S, Germany, France, or India. Twitter is a major platform where we observe these trends. In this work, we examine the magnitude of political discussions on twitter by contrasting the platform usage on levels like gender, political party, and geography, in 2014 and 2019 Indian General Elections. In a

debates have moved from televisions and newspapers to Online Social Networks (OSNs). A variety of research communities from sociology to social network security are showing deep interest in these trends [7]. Changing patterns of user behavior and collective impacts of these changes are being studied in extensive depth [12]. While every aspect of human interaction is essential, elections are of particular interest. The election results impact a nation's future on various fronts like finance, defence, foreign relations, including the day-to-day life of a citizen. India has seen the biggest elections

Note: Urbanization and Literacy as factors in Politicians' Social Media Use in a largely Rural State: Evidence from Uttar Pradesh, India

Asmit Kumar Singh*
asmit18025@iiitd.ac.in
Indraprastha Institute of Information Technology, Delhi
Delhi, Delhi, India

Jivitesh Jain*
jivitesh.jain@students.iiit.ac.in
International Institute of Information Technology, Hyderabad
Hyderabad, Telangana, India

Lalitha Kameswari*
v.a.lalitha@research.iiit.ac.in
International Institute of Information Technology, Hyderabad
Hyderabad, Telangana, India

Ponnurangam Kumaraguru
pk.guru@iiit.ac.in
International Institute of Information Technology, Hyderabad
Hyderabad, Telangana, India

Joyojeet Pal
joyojeet@umich.edu
University of Michigan
Ann Arbor, Michigan, USA

ABSTRACT

With Twitter growing as a preferred channel for outreach among major politicians, there have been focused efforts on online communication, even in election campaigns in primarily rural regions. In this paper, we examine the relationship between politicians' use of social media and the level of urbanization and literacy by compiling a comprehensive list of Twitter handles of political party functionaries and election candidates in the run-up to the 2022 State Assembly elections in Uttar Pradesh, India. We find statistically significant relationships between political Twitter presence and levels of urbanization and with levels of literacy. We also find a strong correlation between vote share and Twitter presence in the winning party, a relationship that is even stronger in urban districts. This provides empirical evidence that social media is already a central part of electoral outreach processes in the Global South, but that this is still selectively more relevant to voters in, and politicians standing for elections from urban and higher-educated regions.

CCS CONCEPTS

• Human-centered computing → Empirical studies in collab-

ACM Reference Format:
Asmit Kumar Singh, Jivitesh Jain, Lalitha Kameswari, Ponnurangam Kumaraguru, and Joyojeet Pal. 2022. Note: Urbanization and Literacy as factors in Politicians' Social Media Use in a largely Rural State: Evidence from Uttar Pradesh, India. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3530190.3534845

1 INTRODUCTION

The past decade has seen an astronomical increase in access to the Internet and various social media platforms like Twitter, Facebook, Instagram, etc. This has resulted in the availability of a huge amount of data which allows insight into various aspects of communicative behavior in social and political settings. In politics, online outreach and propaganda, as well as campaigning during elections, has grown in size on account of the convenience of direct politician-to-citizen communications, bypassing professional journalists, as well as the temporal benefits of allowing political actors to communicate when and how they want to. While no research has shown

“I won the election”

“I Won the Election!”: An Empirical Analysis of Soft Moderation Interventions on Twitter

Savvas Zannettou
Max Planck Institute for Informatics
szannett@mpi-inf.mpg.de

Abstract

Over the past few years, there is a heated debate and serious public concerns regarding online content moderation, censorship, and the principle of free speech on the Web. To ease these concerns, social media platforms like Twitter and Facebook refined their content moderation systems to support soft moderation interventions. Soft moderation interventions refer to warning labels attached to potentially questionable or harmful content to inform other users about the content and its nature while the content remains accessible, hence alleviating concerns related to censorship and free speech.

In this work, we perform one of the first empirical studies on soft moderation interventions on Twitter. Using a mixed-methods approach, we study the users who share tweets with warning labels on Twitter and their political leaning, the engagement that these tweets receive, and how users interact with tweets that have warning labels. Among other things, we find that 72% of the tweets with warning labels are shared by Republicans, while only 11% are shared by Democrats. By analyzing content engagement, we find that tweets with warning labels had more engagement compared to tweets without warning labels. Also, we qualitatively analyze how users in-

Figure 1: An example of a soft moderation intervention on Twitter.

moderation should be performed in a timely manner to ensure that harmful content is removed fast and only a small number of users are exposed to harmful content. This is a tough challenge given the scale of modern social media platforms like Twitter and Facebook. Second, content moderation should be consistent and fair across users. Finally, content moderation should ensure that it is in accordance with

https://ojs.aaai.org/index.php/ICWSM/article/view/18110/17913

Where to start if you are interested?

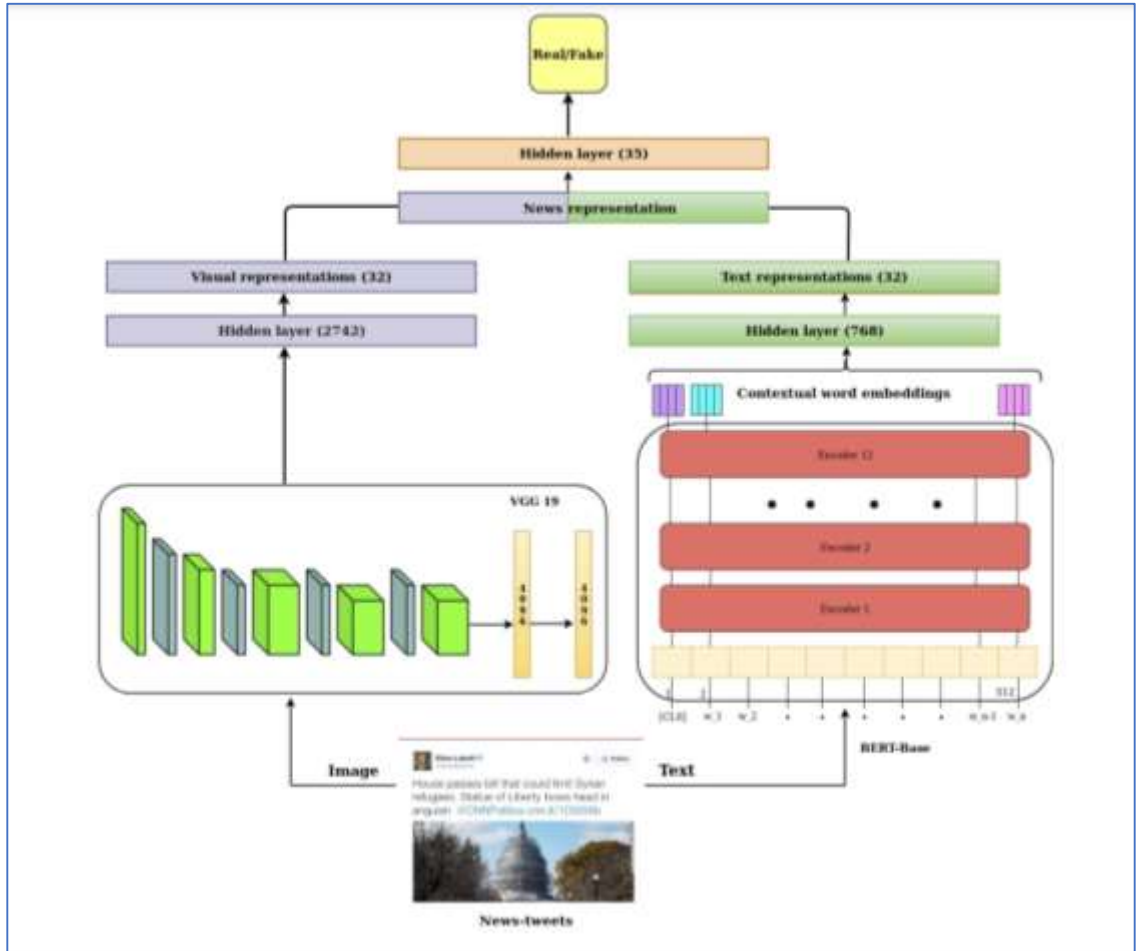
Joyojeet Pal https://joyojeet.people.si.umich.edu/

Ashwin Rajadesingan https://ashwinrajadesingan.com/

Trivedi Centre for Political Data https://tcpd.ashoka.edu.in/

Off course, Precog ☺ https://precog.iiit.ac.in/


Streaming API
Data Collection Tweets Parsing
Data Tagging
CrowdFlower
Feature Extraction
Learning to Rank Build Model
Evaluation
Best Model Selection
tweetcred
System Implementation
User Feedback Usability analysis


Real/Fake
Hidden layer (35)
News representation
Visual representations (32)
Hidden layer (2742)
Text representations (32)
Hidden layer (768)
Contextual word embeddings
VGG 19
Image
Text
BERT-Base
News-tweets

Credibility Ranking of Tweets during High Impact Events

Aditi Gupta, Ponnurangam Kumaraguru
Indraprastha Institute of Information Technology, Delhi, India
{aditig, pk}@iiitd.ac.in
precog.iiitd.edu.in

ABSTRACT

Twitter has evolved from being a conversation or opinion sharing medium among friends into a platform to share and disseminate information about current events. Events in the real world create a corresponding spur of posts (tweets) on Twitter. Not all content posted on Twitter is trustworthy or useful in providing information about the event. In this paper, we analyzed the credibility of information in tweets corresponding to fourteen high impact news events of 2011 around the globe. From the data we analyzed, on average 30% of total tweets posted about an event contained situational information about the event while 14% was spam. Only 17% of the total tweets posted about the event contained situational awareness information that was credible. Using regression analysis, we identified the important content and sourced based features, which can predict the credibility of information in a tweet. Prominent content based features were number of unique characters, swear words, pronouns, and emoticons in a tweet, and user based fea-

1. INTRODUCTION

With the evolution of online social networking and microblogging mediums, two major changes have occurred in the landscape of the Internet usage – firstly, the Internet is replacing traditional media like television and print media as a source for obtaining news and information about current events [16]; secondly, the Internet has provided a platform for common people to share information and express their opinions. Quick response time and high connectivity speed have fueled the propagation and dissemination of information, by users on online social media services like Facebook, Twitter, and YouTube. Work presented in this paper primarily focuses on Twitter; Twitter is a micro-blogging service, which has gained popularity as a major news source and information dissemination agent over last few years. Users on Twitter, create their public / private profile and post messages (also referred as tweets or statuses) via the profile. The maximum length of the tweet can be 140 characters. Each post on Twitter is characterized by two main

$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing Fake Content on Twitter

Aditi Gupta*, Hemank Lamba**, Ponnurangam Kumaraguru*
*Indraprastha Institute of Information Technology, Delhi, India
**IBM Research Labs, Delhi, India
Email: {aditig, pk}@iiitd.ac.in, helamba1@in.ibm.com

***Abstract*—Online social media has emerged as one of the prominent channels for dissemination of information during real world events. Malicious content is posted online during events, which can result in damage, chaos and monetary losses in the real world. We analyzed one such media i.e. Twitter, for content generated during the event of Boston Marathon Blasts, that occurred on April, 15th, 2013. A lot of fake content and malicious profiles originated on Twitter network during this event. The aim of this work is to perform in-depth characterization of what factors influenced in malicious content and profiles becoming viral. Our results showed that 29% of the most viral content on Twitter, during the Boston crisis were rumors and fake content; while 51% was generic opinions and comments; and rest was true information. We found that large number of users with high social reputation and verified accounts were responsible for spreading the fake content. Next, we used regression prediction model, to verify that, overall impact of all users who propagate the fake content at a given time, can be used to estimate the growth of that content in future. Many malicious accounts were created on Twitter during the Boston event, that were later suspended by Twitter. We identified over six thousand such user profiles, we observed that the creation of such profiles surged considerably**

susceptible to fall for rumors / fake content. In one of the most recent incidents in U.S.A., Dow Jones index plunged 140 points due to a rumor tweet posted from a news agency's (Associated Press) Twitter account [14]. the estimated temporary loss of market cap in the S&P 500 totaled $136.5 billion. The rumor mentioned that U.S.A. president Barack Obama has been injured in twin explosions at the White House. In case of *England Riots*, social media was responsible for spreading and instigating violence amongst people. Many rumors propagated during the riots, which resulted in large scale panic and chaos among the public [34]. Two people were also sentenced for spreading false posts on Facebook during the riots [10]. In another incident in Venezuela, some people had spread rumors on Twitter, to destabilize the banking system of the country [44]. In one of the extreme case, *Twitter terrorists* in Mexico were given thirty years sentence for spreading rumors about a fake shooting by gunmen in schools [1]. As parents rushed to get their children from school, and all telephone lines were jammed, the incorrect information, caused a massive chaos in the city, 26 road accidents, jammed telephone lines

Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy

Aditi Gupta*, Hemank Lamba**, Ponnurangam Kumaraguru*, Anupam Joshi†
*Indraprastha Institute of Information Technology, Delhi, India
**IBM Research Labs, Delhi, India
†University of Maryland Baltimore County, Maryland, USA
{aditig, pk}@iiitd.ac.in, helamba1@in.ibm.com, joshi@cs.umbc.edu

ABSTRACT

In today's world, online social media plays a vital role during real world events, especially crisis events. There are both positive and negative effects of social media coverage of events, it can be used by authorities for effective disaster management or by malicious entities to spread rumors and fake news. The aim of this paper, is to highlight the role of Twitter, during Hurricane Sandy (2012) to spread fake images about the disaster. We identified 10,350 unique tweets containing fake images that were circulated on Twitter, during Hurricane Sandy. We performed a characterization analysis, to understand the temporal, social reputation and influence patterns for the spread of fake images. Eighty six percent of tweets spreading the fake images were retweets, hence very few were original tweets. Our results showed that top thirty users out of 10,215 users (0.3%) resulted in 90% of the retweets of fake images; also network links such as follower relationships of Twitter, contributed very less (only

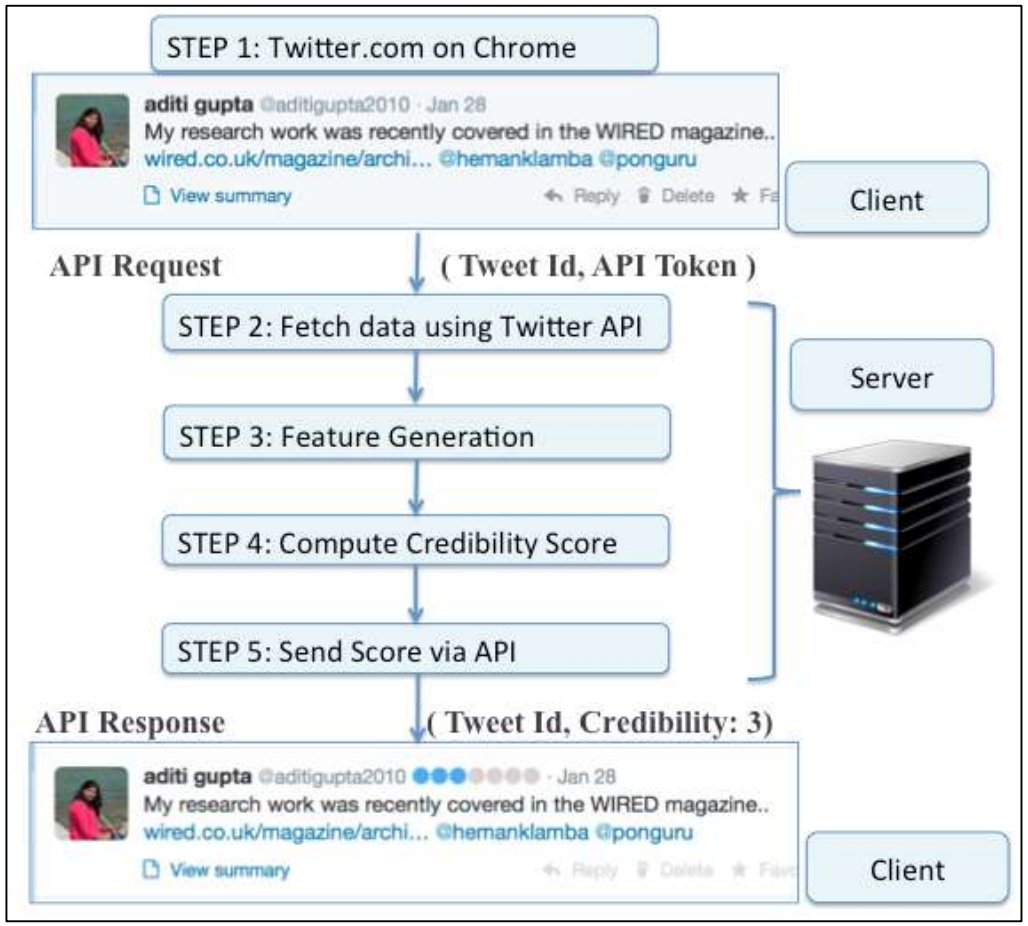
content is posted on OSM, not all of the information is of good quality with respect to the event, like it may be fake, incorrect or noisy. Extracting good quality information is one of the biggest challenges in utilizing information from OSM. Over last few years, people have highlighted how OSM can be used to help in extracting useful information about real life events. But, on the other hand, there have been many instances which have highlighted the negative effects on content on online social media on real life events. The information shared and accessed on social media such as Twitter, is in real-time, the impact of any malicious intended activity, like spreading fake images and rumors needs to be detected and curbed from spreading immediately. Such false and incorrect information can lead to chaos and panic among people on the ground. Since detecting whether images posted are fake or not, using traditional image analysis methods, can be highly time and resource consuming, we explore the option of using Twitter specific features, like the content of the tweet and the user details, in identifying fake images

TweetCred: Real-Time Credibility Assessment of Content on Twitter

Aditi Gupta[1], Ponnurangam Kumaraguru[1], Carlos Castillo[2], and Patrick Meier[2]

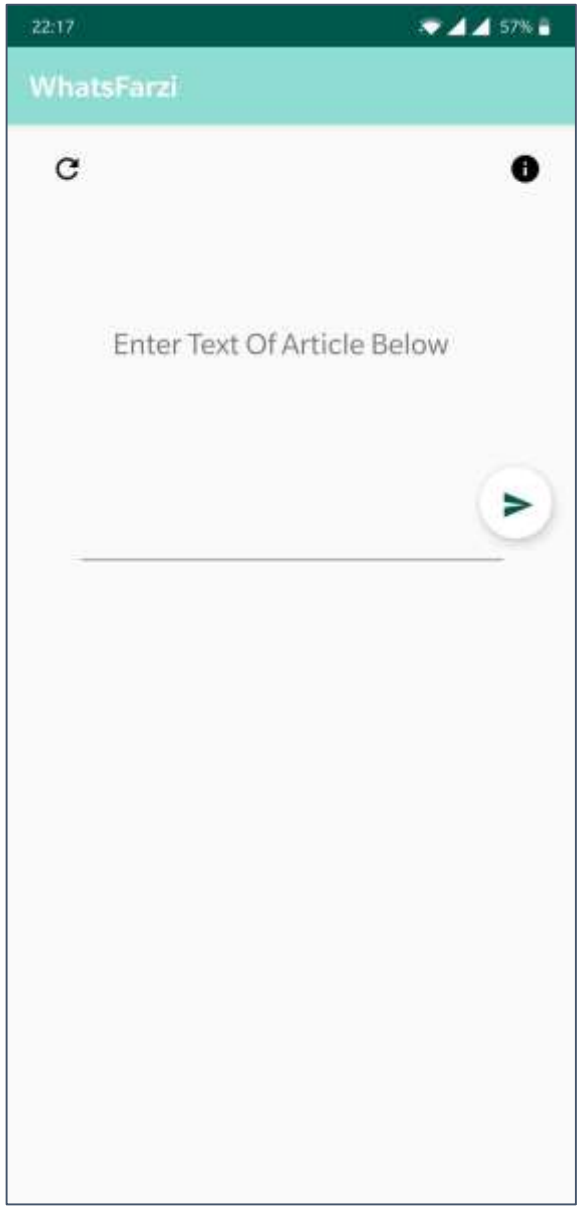
[1] Indraprastha Institute of Information Technology, Delhi, India
{aditig, pk}@iiitd.ac.in
[2] Qatar Computing Research Institute, Doha, Qatar
chato@acm.org, pmeier@qf.org.qa

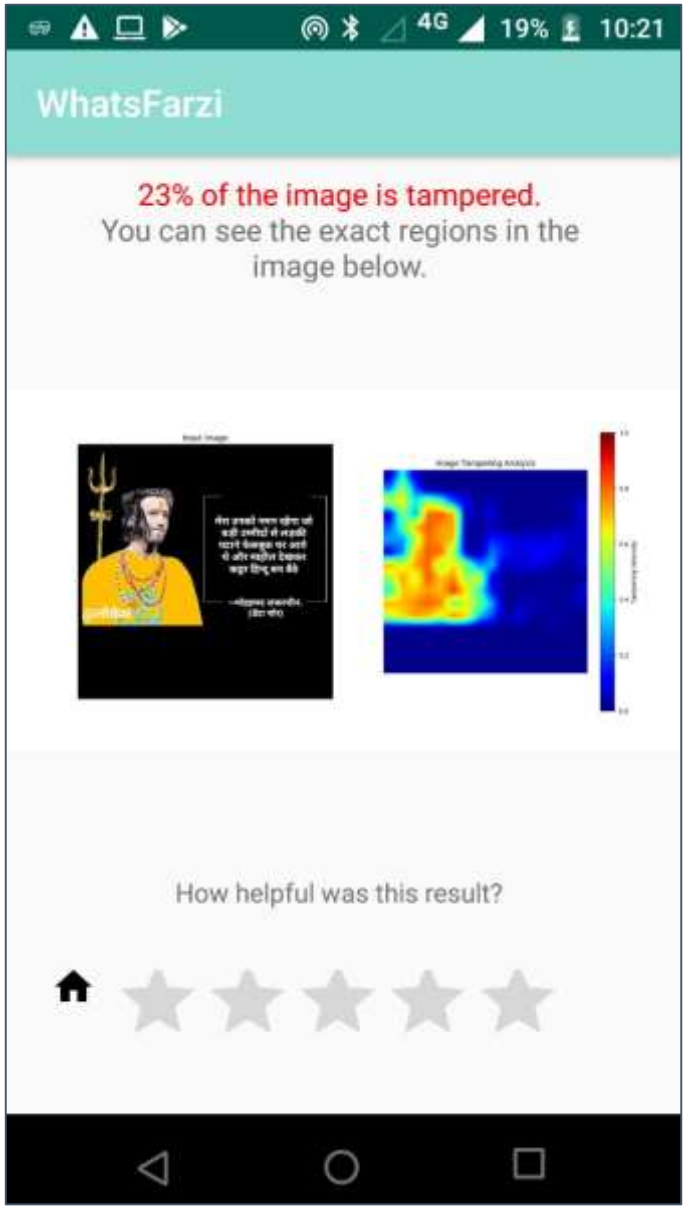
Abstract. During sudden onset crisis events, the presence of spam, rumors and fake content on Twitter reduces the value of information contained on its messages (or "tweets"). A possible solution to this problem is to use machine learning to automatically evaluate the credibility of a tweet, i.e. whether a person would deem the tweet believable or trustwor-


STEP 1: Twitter.com on Chrome
aditi gupta @aditigupta2010 · Jan 28
My research work was recently covered in the WIRED magazine..
wired.co.uk/magazine/archi… @hemanklamba @ponguru
View summary
Client
API Request
(Tweet Id, API Token)
STEP 2: Fetch data using Twitter API
Server
STEP 3: Feature Generation
STEP 4: Compute Credibility Score
STEP 5: Send Score via API
API Response
(Tweet Id, Credibility: 3)
aditi gupta @aditigupta2010 · Jan 28
My research work was recently covered in the WIRED magazine..
wired.co.uk/magazine/archi… @hemanklamba @ponguru
View summary
Client


that's life!
16 hrs
Like Page
We couldn't believe this awful news!
Reports of Rowan Atkinson's death devastate Mr Bean fans around the world
Say it's not true!
THATSLIFE.COM.AU
40
32 Comments
21 Shares
Like
Comment
Share

WhatsFarzi





WhatsFarzi

SpotFake: A Multi-modal Framework for Fake News Detection

Shivangi Singhal
IIIT-Delhi
Delhi,India
shivangis@iiitd.ac.in

Rajiv Ratn Shah
IIIT-Delhi
Delhi, India
rajivratn@iiitd.ac.in

Tanmoy Chakraborty
IIIT-Delhi
Delhi, India
tanmoy@iiitd.ac.in

Ponnurangam Kumaraguru
IIIT-Delhi
Delhi, India
pk@iiitd.ac.in

Shin'ichi Satoh
NII
Tokyo, Japan
satoh@nii.ac.jp

***Abstract*—A rapid growth in the amount of fake news on social media is a very serious concern in our society. It is usually created by manipulating images, text, audio, and videos. This indicates that there is a need of multimodal system for fake news detection. Though, there are multimodal fake news detection systems but they tend to solve the problem of fake news by considering an additional sub-task like event discriminator and finding correlations across the modalities. The results of fake news detection are heavily dependent on the subtask and in absence of subtask training, the performance of fake news detection degrade by 10% on an average.**

To solve this issue, we introduce SpotFake- a multi-modal framework for fake news detection. Our proposed solution detects fake news without taking into account any other subtasks. It exploits both the textual and visual features of an article. Specifically, we made use of language models (like BERT) to learn text features, and image features are learned from VGG-19 pre-trained on ImageNet dataset. All the experiments are performed on two publicly available datasets, *i.e.*, Twitter and Weibo. The proposed model performs better than the current state-of-the-art on Twitter and Weibo datasets by 3.27% and 6.83%, respectively.

***Keywords*-Fake News Detection, Multimedia, Social Computing, Natural Language Processing, Deep Learning**

I. Introduction

"Fake news" is not new. The root of it existed in the soci-

Figure 1: An example of fake news that claims that the actor Sylvester Stallone died due to prostate cancer.

[4], [5] and the authenticity of the news being read. This can have a huge negative impact on the masses. The second reason is lack of automated fact checking methods. Websites such as Politifact[1], Full Fact[2] and AltNews[3] make efforts in fake news detection but the time consuming manual method is too slow to prevent initial spread of a fake news.
This may be due to the fact that textual claims are not

The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning (Student Abstract)

Shivangi Singhal,[1] Anubha Kabra,[2*] Mohit Sharma,[3*]
Rajiv Ratn Shah,[4] Tanmoy Chakraborty,[5] Ponnurangam Kumaraguru[6]
[1,3,4,5,6]IIIT-Delhi, India, [2]DTU, India
(shivangis[1], rajivratn[4], tanmoy[5], pk[6])@iiitd.ac.in, {anubhakabradtu, mohit.sharma.cs29}@gmail.com[2,3]

Abstract

In recent years, there has been a substantial rise in the consumption of news via online platforms. The ease of publication and lack of editorial rigour in some of these platforms have further led to the proliferation of fake news. In this paper, we study the problem of detecting fake news on the FakeNewsNet repository, a collection of full length articles along with associated images. We present SpotFake+, a multimodal approach that leverages transfer learning to capture semantic and contextual information from the news articles and its associated images and achieves the better accuracy for fake news detection. To the best of our knowledge, this is the first work that performs a multimodal approach for fake news detection on a dataset that consists of full length articles. It outperforms the performance shown by both single modality and multiple-modality models. We also release the pretrained model for the benefit of the community.

capture the features from the image modality that may seek to emphasize certain facts.

To overcome the above mentioned challenges, we propose SpotFake+, an advanced version of existing multimodal fake news detection system, named SpotFake (Singhal et al. 2019). The proposed architecture leverages pretrained language transformers and pre-trained ImageNet models for feature extraction. These feature vectors are fed into fully connected layers for classification.

Methodology

Our methodology primarily consists of the following: (i) details of the dataset used in experiments, (ii) pre-processing of the dataset, and (iii) details of the text and image sub-module used in SpotFake+ for fake news detection.

Dataset

https://precog.iiit.ac.in/pubs/SpotFake-IEEE_BigMM.pdf https://precog.iiit.ac.in/pubs/SpotFake_plus_AAAI.pdf
https://precog.iiit.ac.in/pubs/FactDrill_ICWSM2022_final_version.pdf

Where to start if you are interested?

David G Rand https://davidrand-cooperation.com/

Gordon Pennycook https://gordonpennycook.com/

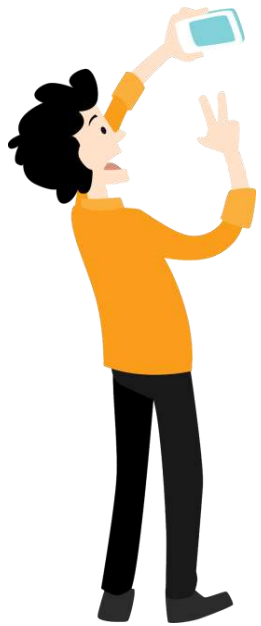
Neelanjan Sircar https://twitter.com/neelanjansircar

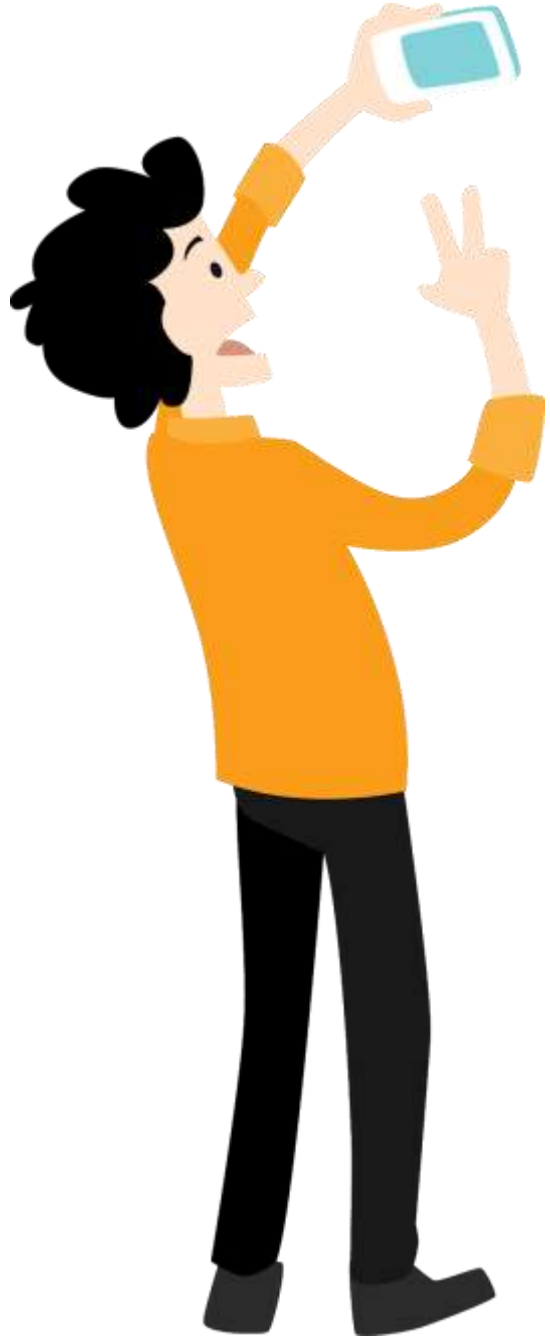
Sumitra Badrinathan https://sumitrabadrinathan.github.io/

Off course, Precog ☺ https://precog.iiit.ac.in/

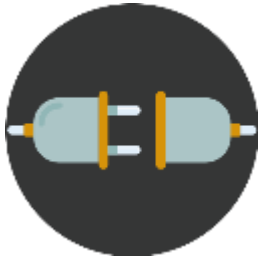
Selfie

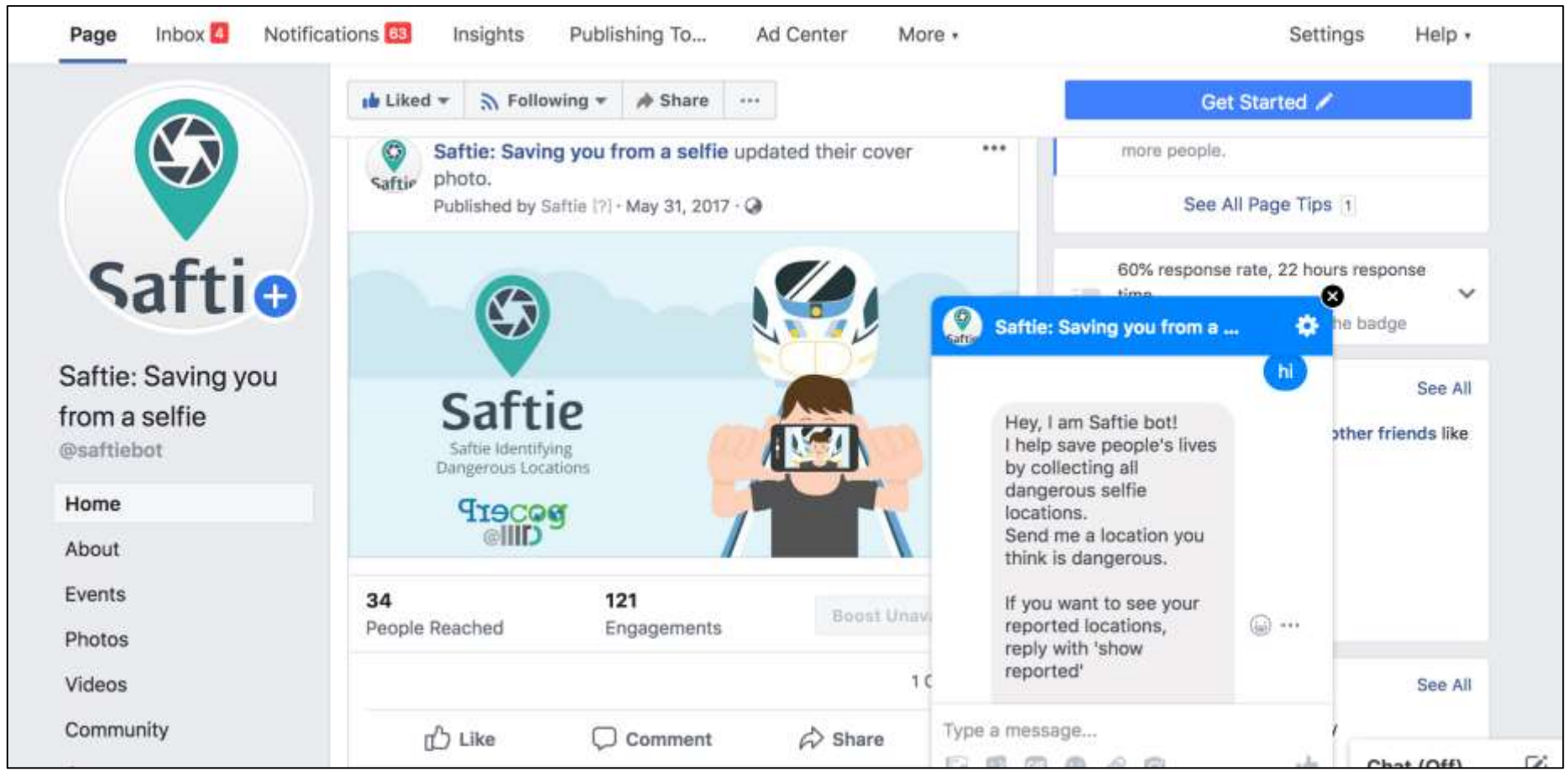

Ellen DeGeneres
@TheEllenShow
Follow
If only Bradley's arm was longer.
Best photo ever. #oscars
pic.twitter.com/C9U5NOtGap
Reply
Retweet
Favorite
More
RETWEETS
2,171,0
FAVORITES
1,048,8
8:36 AM - 3 Mar 2014
Flag media





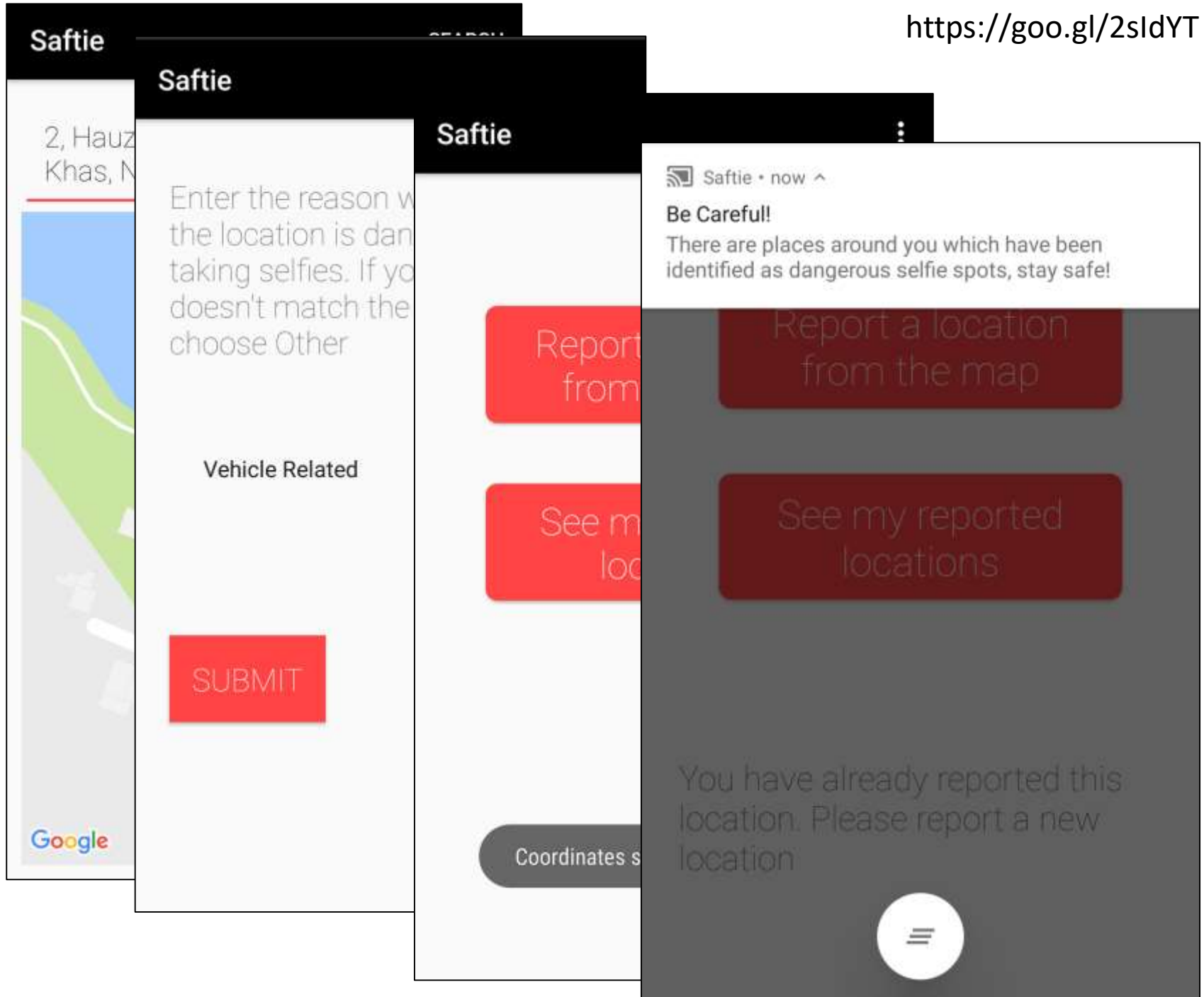


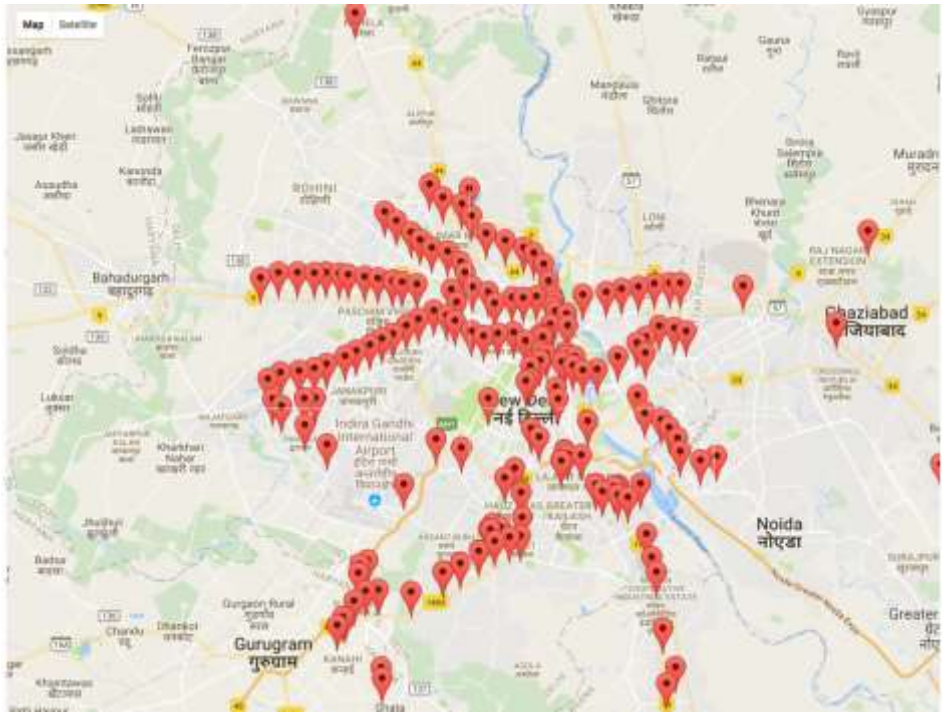



https://www.facebook.com/saftiebot/

https://goo.gl/2sIdYT

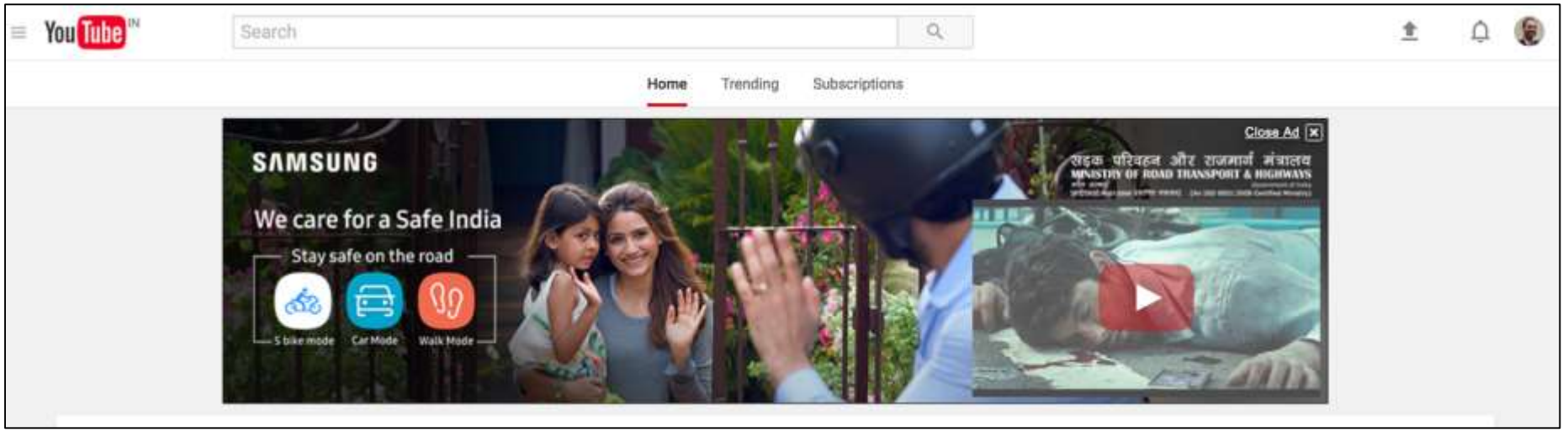


2,000+ location, 400+ verified



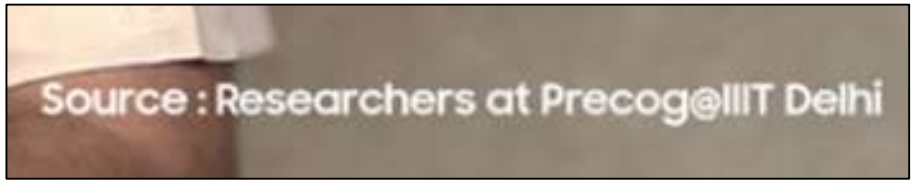
http://precog.iiitd.edu.in/blog/2016/11/killfie-journey/

Samsung #SafeIndia



Samsung cares for a Safe India

150,293,765 views



#Netflix #PatriotAct #HasanMinhaj

Immigration Enforcement | Patriot Act with Hasan Minhaj | Netflix

634,068 views

634,068 views

15K 775 SHARE SAVE ...

http://bit.ly/saftie-cam

Stop the KillFies! Using Deep Learning Models to Identify Dangerous Selfies

Vedant Nanda*, Hemank Lamba§, Divyansh Agarwal*, Megha Arora§, Niharika Sachdeva*, Ponnurangam Kumaraguru*

*IIIT-Delhi, § Carnegie Mellon University
{vedant15114,divyansha,niharikas,pk}@iiitd.ac.in,{hlamba}@cs.cmu.edu,{marora}@andrew.cmu.edu

ABSTRACT

Selfies have become a prominent medium for self-portrayal on social media. Unfortunately, certain social media users go to extreme lengths to click selfies, which puts their lives at risk. Two hundred and sixteen individuals have died since March 2014 until January 2018 while trying to click selfies. It is imperative to be able to identify dangerous selfies posted on social media platforms to be able to build an intervention for users going to extreme lengths for clicking such selfies. In this work, we propose a convolutional neural network based classifier to identify dangerous selfies posted on social media using only the image (no metadata). We show that our proposed approach gives an accuracy of 98% and performs better than previous methods.

ACM Reference Format:
Vedant Nanda*, Hemank Lamba§, Divyansh Agarwal*, Megha Arora§, Niharika Sachdeva*, Ponnurangam Kumaraguru*. 2018. Stop the KillFies! Using Deep Learning Models to Identify Dangerous Selfies. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3184558.3191575

1 INTRODUCTION

A selfie is defined as *a photograph that one has taken of oneself, typically taken with a smartphone or a webcam and shared via social media* [25]. The popularity of selfie culture can be estimated from the fact that in 2015, 24 billion selfies were uploaded to Google

engage in such situations to portray themselves as adventurous and enhance their appearance to others while risking their own physical well-being [9, 18]. Continuing the statistic in [16], we found that as many as 216 individuals have died while attempting to take selfies.

We define a dangerous selfie as *a selfie which potentially might cause harm to an individual or a group that may occur while the individual(s) attempts to take a selfie.* To be able to detect the users who post such dangerous selfies, and to make an intervention, it is essential to find and identify dangerous selfies. By identifying such selfies being posted on the social media platform by a user, combined with the frequency at which the user is posting them, the social networking platform can decide if a particular user is overindulging in risk-taking behavior, which could potentially be harmful to their health. In this work, we propose a deep-learning based framework to identify dangerous selfies posted on Twitter. We use existing deep neural networks such as VGG16 and VGG19 [20], Inception v3 [23], ResNet50 [11] etc. and adapt it to perform well on the task of detecting dangerous selfies. We discover that our model outperforms the previously proposed models by a factor of 1.34 in terms of accuracy on the test set. We believe that this work will help researchers understand a user's propensity to post such selfies on online social media in a much better way, thus resulting in effective intervention technologies.

2 RELATED WORK

From Camera to Deathbed: Understanding Dangerous Selfies on Social Media

Hemank Lamba,[1] Varun Bharadhwaj,[3] Mayank Vachher,[2] Divyansh Agarwal,[2] Megha Arora,[1] Niharika Sachdeva,[2] Ponnurangam Kumaraguru[2]
[1]Carnegie Mellon University, USA, [2]Indraprastha Institute of Information Technology, Delhi, India
{hlamba@cs,marora@andrew}.cmu.edu, {mayank13059,divyansha,niharikas,pk}@iiitd.ac.in
[3]National Institute of Technology, Tiruchirappalli
var6595@gmail.com

Abstract

Selfie culture has emerged as a ubiquitous instrument for self-portrayal in recent years. To portray themselves differently and attractive to others, individuals may risk their life by clicking selfies in dangerous situations. Consequently, selfies have claimed 137 lives around the world since March 2014 until December 2016. In this work, we perform a comprehensive analysis of the reported selfie-casualties and note various reasons behind these deaths. We perform an in-depth analysis of such selfies posted on social media to identify dangerous selfies and explore a series of statistical models to predict dangerous posts. We find that our multimodal classifier using combination of text-based, image-based and location-based features performs the best in spotting dangerous selfies. Our classifier is trained on 6K annotated selfies collected on Twitter and gives 82% accuracy for identifying whether a selfie posted on Twitter is dangerous or not.

Introduction

A selfie is defined as *a photograph that one has taken of oneself, typically taken with a smartphone or webcam and shared via social media* (Taslim and Rezwan 2013). In 2015, Google estimated that 24 billion selfies were uploaded to Google Photos [1] and the number of selfies posted on Instagram increased by 900 times between 2012 and 2014 (Souza

dicating the dangers of taking selfies,[2] and Indian authorities including Mumbai police and Indian Railways issued warning for taking selfies at dangerous locations.[3]

Despite the increase in incidents where selfies were the reason behind physical harm caused to individuals, few research works explore factors that may result into dangerous selfies. Studies have indicated clicking selfies at dangerous locations as one of the reasons for selfie-related casualties (Bhogesha, John, and Tripathy 2016). Social media has emerged as a powerful medium to share and gain attention through such dangerous selfies (Souza et al. 2015). Given the popularity of the selfie culture, and increasing number of selfie deaths, it is crucial to characterize and predict the behavior of taking/posting dangerous selfies on social media. However, this remains largely unexplored. In this work, we try to identify features that can be derived from selfies posted on social media to predict dangerous selfies.

We formulate our specific research goals as:

1. Analyze incidents associated with reported fatal selfie casualties to understand the reasons behind the deaths and characterize such selfie-ers.
2. Investigate the content measures derived from social media that are predictive of dangerous selfies.

In this paper, we first comprehensively analyze the deaths

https://precog.iiit.ac.in/pubs/camera-to-deathbed-icwsm2017.pdf ICWSM 2017

Are you a Spotify user?

We are looking for people* to participate in a study that associates Music consumption with well-being

What to do?

1. Request your Spotify listening history using given steps

2. Fill the survey form

*Must be an active spotify user for atleast a year

Monetary compensation of ₹200 offered for your time & effort!

Prof. Vinoo Alluri
Prof. Ponnurangam Kumaraguru (PK)
Dr. Geetika Arora

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Takeaways

No dearth of problems to study…

Many problems are resource poor, e.g. code mix is low resource, hard to collect corpus while being abundant in social networks ….

Social impact & utility of our work

Will be happy to discuss anything further…

Looking for (including PhD) students / RAs …

is a group of researchers who study, analyze, and build different aspects of social systems (e.g. social web systems like Twitter, Facebook), including their security and privacy. By understanding and measuring complex networks, we try and build solutions for social good. Our work primarily derives from Data Science, Computational Social Science, Social Computing, Machine Learning, and Natural Language Processing.

https://precog.iiit.ac.in/

Join us in solving interesting problems!

https://forms.office.com/r/bQ7SX90pZL

Connect with us on Social!

https://forms.gle/HW2b3kpsXweuD3w98

https://precog.iiit.ac.in/pages/pk-social

Thanks!

Questions?


pk.profgiri


linkedin/in/ponguru


pk.guru@iiit.ac.in


http://precog.iiit.ac.in/


@ponguru