

PSOSM

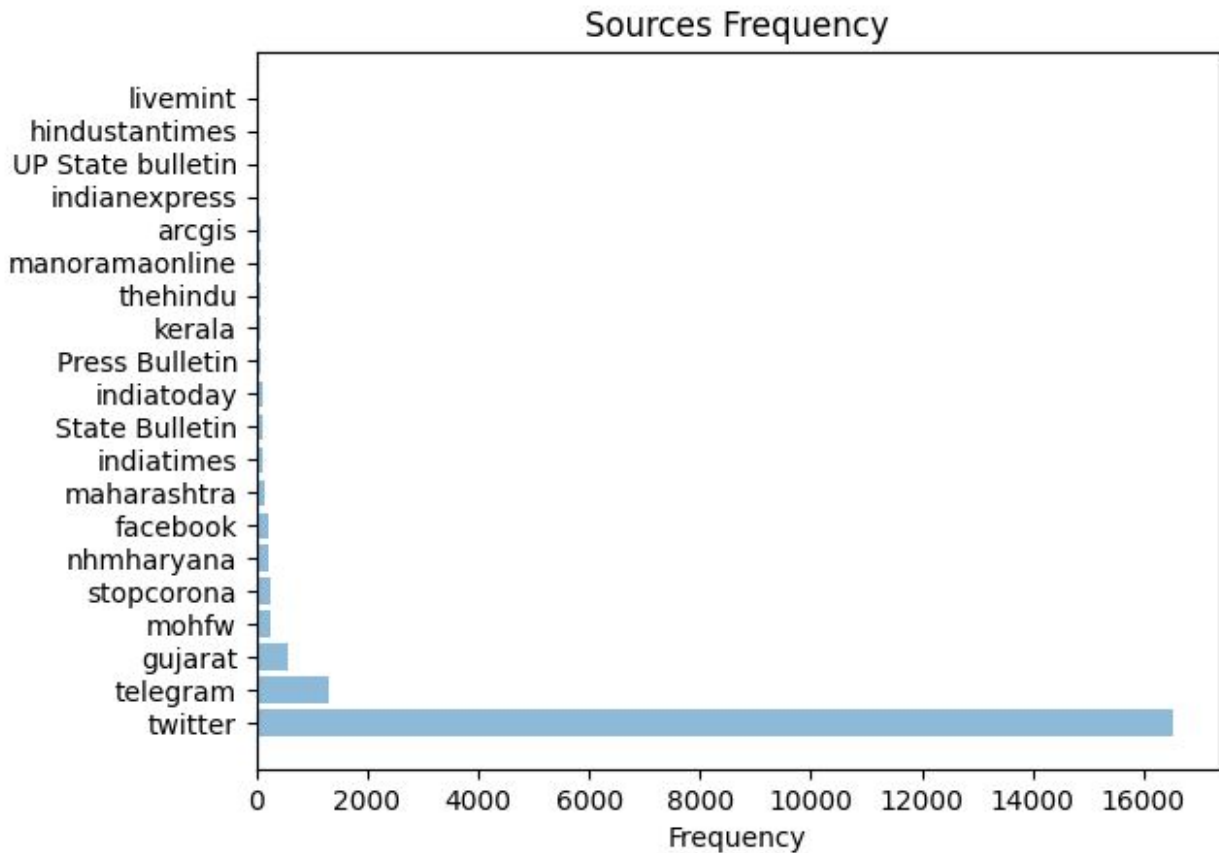
Endsem Take Home

Members

1. Mohit Juneja 2017067
2. Preyansh Rastogi 2017176
3. Aarish Chhabra 2017212

(a) All Sources -

['twitter', 'telegram', 'gujarat', 'mohfw', 'stopcorona', 'nhmharyana', 'facebook', 'maharashtra', 'indiatimes', 'State Bulletin', 'indiatoday', 'Press Bulletin', 'kerala', 'thehindu', 'manoramaonline', 'arcgis', 'indianexpress', 'UP State bulletin', 'hindustantimes', 'livemint', 'dronamaps', 'CM's Meet', 'outlookindia', 'rajswasthya', 'State Bulletin, 6th Apr 8AM', 'deccanherald', 'thenewsminute', 'ndtv', 'punjabupdate', 'newindianexpress', 'ap', 'mathrubhumi', 'State Health Bulletin', 'karnataka', 'State Bulletin, 6th Apr', 'State bulletin', 'google', 'jagran', 'chdcovid19', 'dbpost', 'State Health Bulletin - 23rd Mar 2pm', 'wbhealth', 'gujaratsamachar', 'odisha', 'businesstoday', 'indiatvnews', 'yahoo', 'amarujala', 'cghealth', 'firstpost', 'State Health Bulletin - 24th Mar 8am', 'thewire', 'telangana', 'delhigovt', 'theweek', 'businessinsider', 'news18', 'bhaskar', 'naidunia', 'State Health Bulletin - 22nd March', 'Greater Mumbai Health Department Update', 'patrika', 'ptinews', 'nrhmhp', 'weather', 'tribuneindia', 'UT Health Bulletin', 'business-standard', 'UT Bulletin, 6th Apr upto 3PM', 'youtube', 'theprint', 'varthabharati', 'manoramanews', 'https', 'indiatv', 'thekashmirmonitor', 'stopcoronatn', 'financialexpress', 'jipmer', 'py', 'newsonair', 'thepigeonexpress', 'latestly', 'kalingatv', 'P62929', 'babushahi', 'asianetnews', 'twimg', 'oneindia', 'punemirror check for recovery news', 'P56', 'economictimes', 'Gujarat Dashboard', 'kannadaprabha', 'sakaltimes', 'ladakh']



Considering only top 30 srcs in graph(viewable)

(b)

Twitter is the major information source compared to other sources.

Twitter has contributed 16538 times.

However, the second most contribution is done by Gujarat News and it has contributed 574 times.

(c)

10 most recent tweets

1. #BiharFightsCorona fifth update of the day.3 more covid-19 positive cases in bihar taking the total to 96. 3-male 30,36 and 52 years from jamalpur ,munger. came in touch with covid-19 positive. contact tracing is on.
2. #BiharFightsCorona fifth update of the day.3 more covid-19 positive cases in bihar taking the total to 96. 3-male 30,36 and 52 years from jamalpur ,munger. came in touch with covid-19 positive. contact tracing is on.
3. #BiharFightsCorona fifth update of the day.3 more covid-19 positive cases in bihar taking the total to 96. 3-male 30,36 and 52 years from jamalpur ,munger. came in touch with covid-19 positive. contact tracing is on.

Most prominent words are

1. Positive case
2. Health Department
3. COVID 19
4. Maharashtra
5. Total
6. Number
7. Positive
8. DelhiFightsCorona
9. Health bulletin

(e)

Tweets from Verified Sources: 14714

Tweets from Non-Verified Sources: 973

We have defined our own criteria for giving a **credibility score** to non-verified sources.

This takes into account the following 4 parameters:

- Follower Count / Friends Count > 20
- Number of Posts > 200
- Has Description
- Has account created in previous year, as the first case was observed on 30th jan.

All these parameters result in 1/0 and the summation of these gives a **credibility score** in the **range of 1-4** in the **increasing order** of credibility.

Count of users getting a score 4: 408

Count of users getting a score 3: 564

Count of users getting a score 2: 0

Count of users getting a score 1: 1

Count of users getting a score 0: 0

Sample Non-Verified Sources with the Credibility scores:

[Rohit Kansal](#) **Credibility Score: 4**, We should “**believe**” the tweets from this user.

Observations:

Parameter 1 passes (defined above).

Parameter 2 passes (defined above).

Parameter 3 passes (defined above).

Parameter 4 passes (defined above).

Has a government url mentioned in his bio. <http://new.jkdirinf.in/>

[GujHFWDept](#) **Credibility Score: 4**, We should “**believe**” the tweets from this user.

Observations:

Parameter 1 passes (defined above).

Parameter 2 passes (defined above).
Parameter 3 passes (defined above).
Parameter 4 passes (defined above).
Has a government url mentioned in his bio. <https://gujhealth.gujarat.gov.in/>

[TNIE Andhra Pradesh](#) **Credibility Score: 3**, We should “**believe**” the tweets from this user.

Observations:

Parameter 1 fails (defined above).
Parameter 2 passes (defined above).
Parameter 3 passes (defined above).
Parameter 4 passes (defined above).
It is an official account of News Source of Andhra Pradesh and even has it's url in the bio.
In Spite of Parameter 1 failing, we should believe because the reason for this failure is that it is a comparatively new account for an old news channel TNIE operating all across india and even TNIE has a verified account on twitter.

[KBS Sidhu, IAS](#) **Credibility Score: 3**, We should “**believe**” the tweets from this user.

Observations:

Parameter 1 fails (defined above).
Parameter 2 passes (defined above).
Parameter 3 passes (defined above).
Parameter 4 passes (defined above).
Is an IAS officer in Chandigarh, India.
In spite of the failure of parameter 1, the account should be believed as the user is an IAS officer & has number of followers : 25.4 K and number of friends: 1728, ratio: close to 15, and a very old account, which proves that the account is not a fake account.

[Collector DNH](#) **Credibility Score: 1**, We should “**not believe**” the tweets from this user.

Observations:

Parameter 1 passes (defined above).
Parameter 2 fails (defined above).
Parameter 3 fails (defined above).
Parameter 4 fails (defined above).
No Profile Picture

This could be the account of the collector of Dadar and Nagar Haveli. However, on checking the official website of the collectorate of Dadar and Nagar Haveli, we didn't find any mention of the twitter account. There is a possibility that the website is not updated. But, still we should not believe the user as the score is just 1 & with ratio>20 and account being very new, there are higher chances of it being a **bot account**.

(f)

The Major PII Information revealed are

1. Location

Examples: (Some of the outputs from NER spacy using “GPE” label)

[('Pune', 'GPE')]

Public Health Department, Maharashtra: 5 more people found positive in Pimpri-Chinchwad today taking total tally of positive cases in Pune to 15. Maharashtra state tally reaches 31. #Coronavirus

[('Srinagar', 'GPE')]

Rohit Kansal, Principal Secretary Planning, Jammu & Kashmir: First positive case in Kashmir-Khanyar, Srinagar. The patient has history of foreign travel, arrived on 16 March; has been put in isolation. (File pic) <https://t.co/zGdZCmcWvg>

[('Rajasthan', 'GPE')]

Rajasthan: 3 people have tested positive for #Coronavirus, in Jhunjhunu today. <https://t.co/YgLQ4CKrNP>

[('Delhi', 'GPE')]

Total number of coronavirus cases in national capital rises to 14: Delhi health department

[('Jabalpur', 'GPE')]

Madhya Pradesh sees first Covid-19 cases; four test positive for novel coronavirus in Jabalpur

2. Nationality

Examples: (Some of the outputs using NER spacy “NORP” label)

[('Indonesian', 'NORP')]

#update: 5 news cases of #COVID19 in TN. 4 Indonesian nationals & their travel guide from Chennai test positive at #Salem Medical College. Quarantined since 22.3.20 @MoHFW_INDIA @CMOTamilNadu #vijayabaskar

[('Thai', 'NORP')]

#UPDATE: 3 new positive cases of #covid19 in TN. 18 Y M contact of 2nd Pt at #RGGH. 63 Y M Dubai return at #Walajah GH. 66 Y M contact of Thai nationals at #IRTT, Perundurai. Pts are in isolation & stable. @MoHFW_INDIA #TNHealth @CMOTamilNadu #Vijayabaskar

[('Tibetan', 'NORP')]

A 69-year-old Tibetan refugee, who died at a hospital in Tanda in Himachal Pradesh earlier today, has been tested positive for #coronavirus. He returned from the US on March 15: State Additional Chief Secretary (Health) RD Dhiman <https://t.co/BwGfUzII02>

[('Indian', 'NORP')]

Army sources: First positive case of #COVID19 confirmed of an Indian Army jawan (from Ladakh Scouts). Jawan's father has travel history to Iran. Jawan is being treated while his family including sister&wife have been put in quarantine. Jawan's father has also tested positive.

[('Italian', 'NORP')]

69-year-old Italian tourist, who had recovered from COVID-19, dies of cardiac arrest in Jaipur private hospital: SMS Hospital official

3. Travel History

We found this PII using regex pattern matching. We used the following regex.

"(\btravel history to [a-zA-Z]*\b)|(\btravel history to the [a-zA-Z]*\b)|(\breturned from [#a-zA-Z]*\b)|(\bvisited [#a-zA-Z]*\b)"

Eg. #CoronaVirusUpdate: The 27-year-old person from #Nellore, who **returned from #Italy** and was tested positive for #COVID-19, discharged from #Nellore GGH after his recovery @NewIndianXpress

Eg. Maharashtra Health Department: A 59-year-old woman has tested positive for #Coronavirus, she had a **travel history to Russia** and Kazakhstan.

The number of positive cases in the state now rises to 32.
(ANI)

#CautionYesPanicNo

#CoronaVirusUpdates
#COVID19
#Coronaindia <https://t.co/HmsEGla9PY>

Eg. Yavatmal District Collector MD Singh: One more person, with **travel history to Dubai**, has been tested positive for #coronavirus. With this, total number of COVID-19 cases rises to 38 in #Maharashtra. <https://t.co/1zaQS6ixHN> <https://t.co/7m1JdKEg6d>

Eg. Rohit Kansal, Principal Secretary Planning, Jammu & Kashmir: One more case,

with **travel history to Saudi** Arabia, confirmed positive. Had reported high viral load earlier. Total 3 positive cases now in the union territory. #Coronavirus (file pic)
<https://t.co/y1tcFSBbjG>

Eg. Public Health Department, Maharashtra: A 49-year-old person has also been tested positive for #COVID19 in Mumbai today; he has **travel history to USA**. With this, total number of COVID19 cases rises to 41 in #Maharashtra <https://t.co/xNJWxeLouA>

4. Age

We found this PII using regex pattern matching. We used the following regex.

“(\\b\\d*-year-old\\b)|(\\b\\d*-yr-old\\b)|(\\b\\d*-year old\\b)|(\\b\\d* year old\\b)|(\\b\\d*-yr old\\b)|(\\b\\d* yr old\\b)”

Eg. We have got 2 more #COVID2019 cases in Karnataka taking the total number of confirmed cases to 10. **20 yr old** female who travelled from UK is tested positive & another contact of P6 (kalburgi deceased patient) is tested positive. Both are admitted in designated isolation hospital

Eg. Maharashtra Health Department: A **59-year-old** woman has tested positive for #Coronavirus, she had a travel history to Russia and Kazakhstan.

The number of positive cases in the state now rises to 32.
(ANI)

Eg. Sharat B, Deputy Commissioner, Kalaburagi: A **63-year-old** doctor who treated the **76-yr-old** man who died due to #Coronavirus, has tested positive. He along with his family has been kept in quarantine at his home. He will be sent to isolation ward today. #Karnataka

#CautionYesPanicNo

#CoronaVirusUpdates

#COVID19

#Coronaindia <https://t.co/HmsEGla9PY>

Eg. Puducherry reports its first confirmed COVID-19 case; **68- year old** woman tests positive in Mahe region: Health department official

Eg. First positive case of #COVID19 confirmed in West Bengal. The **18-year-old** man has travel history to United Kingdom. He is kept in isolation. <https://t.co/qLW0GbvsWn>

5. Gender

We found this PII using regex pattern matching. We used the following regex.

“(\bman\b)|(\bwoman\b)|(\bboy\b)|(\bgirl\b)|(\bfemale\b)|(\bmale\b)|(\bson\b)|(\bdaughter\b)|(\bmother\b)|(\bfather\b)”

Eg. We have got 2 more #COVID2019 cases in Karnataka taking the total number of confirmed cases to 10. 20 yr old **female** who travelled from UK is tested positive & another contact of P6 (kalburgi deceased patient) is tested positive. Both are admitted in designated isolation hospital

Eg. Maharashtra Health Department: A 59-year-old **woman** has tested positive for #Coronavirus, she had a travel history to Russia and Kazakhstan.

The number of positive cases in the state now rises to 32.

(ANI)

Eg. Sharat B, Deputy Commissioner, Kalaburagi: A 63-year-old doctor who treated the 76-yr-old **man** who died due to #Coronavirus, has tested positive. He along with his family has been kept in quarantine at his home. He will be sent to isolation ward today. #Karnataka

#CautionYesPanicNo

#CoronaVirusUpdates

#COVID19

#Coronaindia <https://t.co/HmsEGla9PY>

Eg. Puducherry reports its first confirmed COVID-19 case; 68- year old **woman** tests positive in Mahe region: Health department official

Eg. The first positive case of #COVID19 confirmed in West Bengal. The 18-year-old **man** has travel history to the United Kingdom. He is kept in isolation. <https://t.co/qLW0GbvsWn>

6. Names

Examples: (Some of the outputs after using NER spacy with “**PERSON**” label)

'Rigzin Sampheal', 'PERSON'

Total Six #COVID19 cases in #Ladakh after 3 more positive cases. UT spokesman Commissioner Secretary Rigzin Sampheal says 2 cases are from Leh, 1 from Kargil district. UT administration imposes section 144 in #Kargil as a preventive measure.

[('Eatala Rajendra', 'PERSON')]

Telangana Health Minister Eatala Rajendra: Another #COVID19 positive case confirmed in the state, positive cases toll rises to 5 in the state. The new confirmed patient is an Indonesian not an Indian national; he has a travel history from Delhi to Telangana.

[('Deepak Mhaisekar', 'PERSON')]

Deepak Mhaisekar, Divisional Commissioner of Pune: One person has been tested positive for #COVID19 in Pimpri-Chinchwad today; he has travel history to USA. With this, total number of COVID19 cases rises to 40 in #Maharashtra <https://t.co/idA7jlvWMV>

[('KK Shailaja', 'PERSON')]

Kerala Health Minister KK Shailaja: Two more #COVID—19 cases reported in the state. One UK national, and a doctor - who returned from abroad & tested positive. The total positive cases in the state rises to 21. <https://t.co/aJ0Sgx5J9l>

[('Rohit Kansal', 'PERSON')]

#COVID-19

Jammu & Kashmir

First positive case in Kashmir- Khanyar, Srinagar. History of foreign travel. Arrived on 16/3/2020. Put in Isolation. Surveillance started in 300m area. Request all to cooperate.

Also Immediately self report any symptoms

@diprjk @HealthMedicalE1

7. Designation/ Job Position

Note: This one is done manually.

Eg. Rohit Kansal, Principal Secretary Planning, Jammu & Kashmir: One more case, with travel history to Saudi Arabia, confirmed positive. Had reported high viral load earlier. Total 3 positive cases now in the union territory. #Coronavirus (file pic) <https://t.co/y1tcFSBbjG>

Eg. Yavatmal District Collector MD Singh: One more person, with travel history to Dubai, has been tested positive for #coronavirus. With this, total number of COVID-19 cases rises to 38 in #Maharashtra. <https://t.co/1zaQS6ixHN> <https://t.co/7m1JdKEg6d>

Eg. Kerala Health Minister KK Shailaja: Two more #COVID—19 cases reported in the state. One UK national, and a doctor - who returned from abroad & tested positive. The total positive cases in the state rises to 21. <https://t.co/aJ0Sgx5J9l>

Eg. Total Six #COVID19 cases in #Ladakh after 3 more positive cases. UT spokesman Commissioner Secretary Rigzin Samphel says 2 cases are from Leh, 1 from Kargil district. UT administration imposes section 144 in #Kargil as a preventive measure.

#AIRReport: Yangchan Dolma <https://t.co/Y5lr6cmZnR>

Eg. Till now, 43 positive cases of Coronavirus have been detected in the state: **Gujarat**
Principal Secretary, Health and Family Welfare Jayanti Ravi <https://t.co/bUfwENmxEt>

g)

Count of the tweets which did mention location: **8521**

Count of the tweets whose at least 1 mentioned location matched with any of the location column:

5283

Count of the tweets whose at least 1 mentioned location matches with either the notes column or the backup notes column: **461**

Required Percentage(Location Columns): 62 %

Required Percentage(Notes Columns): 5%

Library Used for the above location extraction from the text: spacy

Spacy Labels used for the same: GPE, ORG, NORP, LOC

Some Samples of detected locations using spacy:

['wuhan university']

Update on #Coronavirus incident: A patient has been tested positive for Novel #Coronavirus infection and is kept in isolation. The infected patient is a student of Wuhan University. We are closely monitoring the situation.

['kerala', 'kasargod']

Kerala Health Minister KK Shailaja: Third case of #coronavirus tested positive in Kerala in Kasargod. <https://t.co/g390gJ98J0>

['italian', 'jaipur', 'sms hospital']

69-year-old Italian tourist, who had recovered from COVID-19, dies of cardiac arrest in Jaipur private hospital: SMS Hospital official

['italy']

రాష్ట్రంలో నిన్న (31.3.2020) సాయంత్రం 9 గంటల తర్వాత నుంచి ఈ రోజు (01.04.2020) ఉదయం 9:00 వరకు 43 కొత్త కొవిడ్-19 పాజిటివ్ కేసులు నమోదయ్యాయి.

వీటిలో కలిపి మొత్తం 87 పాజిటివ్ కేసులు రాష్ట్రము లో నమోదయ్యాయి.

@AndhraPradeshCM @MoHFW_INDIA

#APFightsCoronaVirus #COVID #covidupdate <https://t.co/NgP9hMi8FN>

['italy']

#CoronaVirusUpdate: The 27-year-old person from #Nellore, who returned from #Italy and was tested positive for #COVID-19, discharged from #Nellore GGH after his recovery @NewIndianXpress

['jammu & kashmir nhm', 'jammu', 'jammu', 'kashmir']

Jammu & Kashmir NHM Director Bhupinder Kumar: A fresh positive case of #Coronavirus has emerged from Jammu. The patient is a close contact of the first confirmed case in Jammu and Kashmir. (File photo) <https://t.co/BbaEvlzwaU>

['public health department', 'maharashtra', 'pimpri-chinchwad', 'pune', 'maharashtra']

Public Health Department, Maharashtra: 5 more people found positive in Pimpri-Chinchwad today taking total tally of positive cases in Pune to 15. Maharashtra state tally reaches 31. #Coronavirus

['public health department', 'maharashtra', 'pimpri-chinchwad', 'pune', 'maharashtra']

Public Health Department, Maharashtra: 5 more people found positive in Pimpri-Chinchwad today taking total tally of positive cases in Pune to 15. Maharashtra state tally reaches 31. #Coronavirus

h) ANI has the most number of COVID19 posts.

Yes, it is verified.