



Fake News Detector

Covid19 Hackathon Idea Submission

Submitted By:

Dharmendra Pathak (AIIR05-Droit)

Ph.D Scholar, A80166220005 (Batch 2)

Amity School of Engineering and Technology

Amity University Chhattisgarh, Raipur



Topics to be covered....

- Problem Statement
- Proposed Solution
- Solution Approach
- Solution Workflow
- RUN-TIME Results
- USP of the Solution



Problem Statement

- To come up with solutions by utilizing heatmaps, like geotagging high risk zones and marking those through heatmaps, or in other ways use clusters to markdown the fake news and its origin.



Proposed Solution

- Generation and using GeoTagging for the detection of Covid-19 Fake News from Twitter Data of HOTSPOT States using Deep Learning Framework

Solution Approach

- The proposed solution generates the geotag of the most affected states by means of the timely synchronized and filtered data set collected from trusted government source. Then on the basis of most affected regions, filter the twitter data of Covid-19 and detect the fake news using Gated recurrent unit (GRU) deep learning algorithm.
- Following are the key elements of the idea proposed:

Solution Approach....

1. Authentic Data Set:

- The data set is collected from the government website i.e. <https://www.mohfw.gov.in/> and timely synchronized with it for the generation of real-time maps. The attributes of data set are Name of state, Total Confirmed cases, Cured/Discharged/Migrated, Death and Twitter Data Set is collected from <https://ieee-dataport.org/open-access/corona-virus-covid-19-tweets-dataset>

2. Platform/Framework used:

- The proposed solution implements the libraries/APIs of Deep Learning using Python Programming Language as follows:
 - HTTP web scraper
 - GeoPandas
 - Matplotlib and Seaborn
 - Online GIS Mapping
 - TensorFlow and Keras
 - SkLearn

Solution Workflow

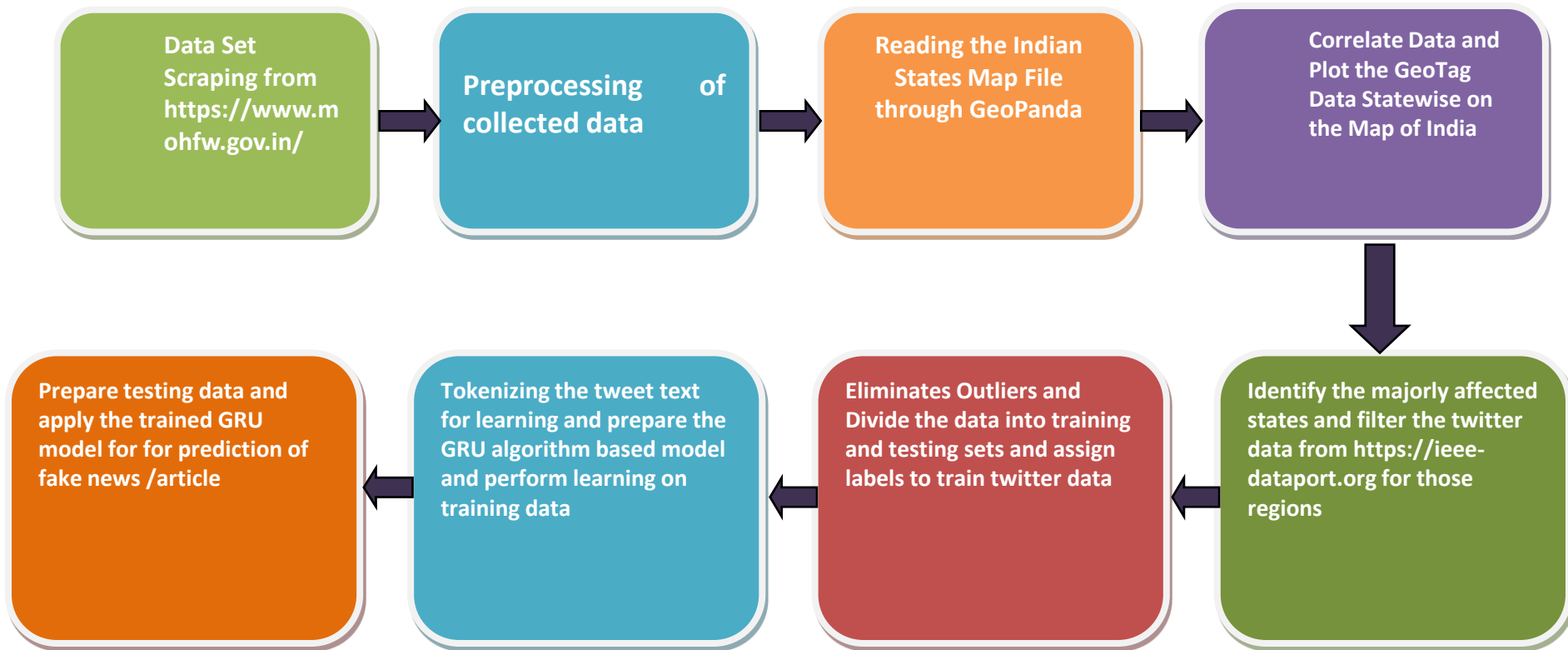
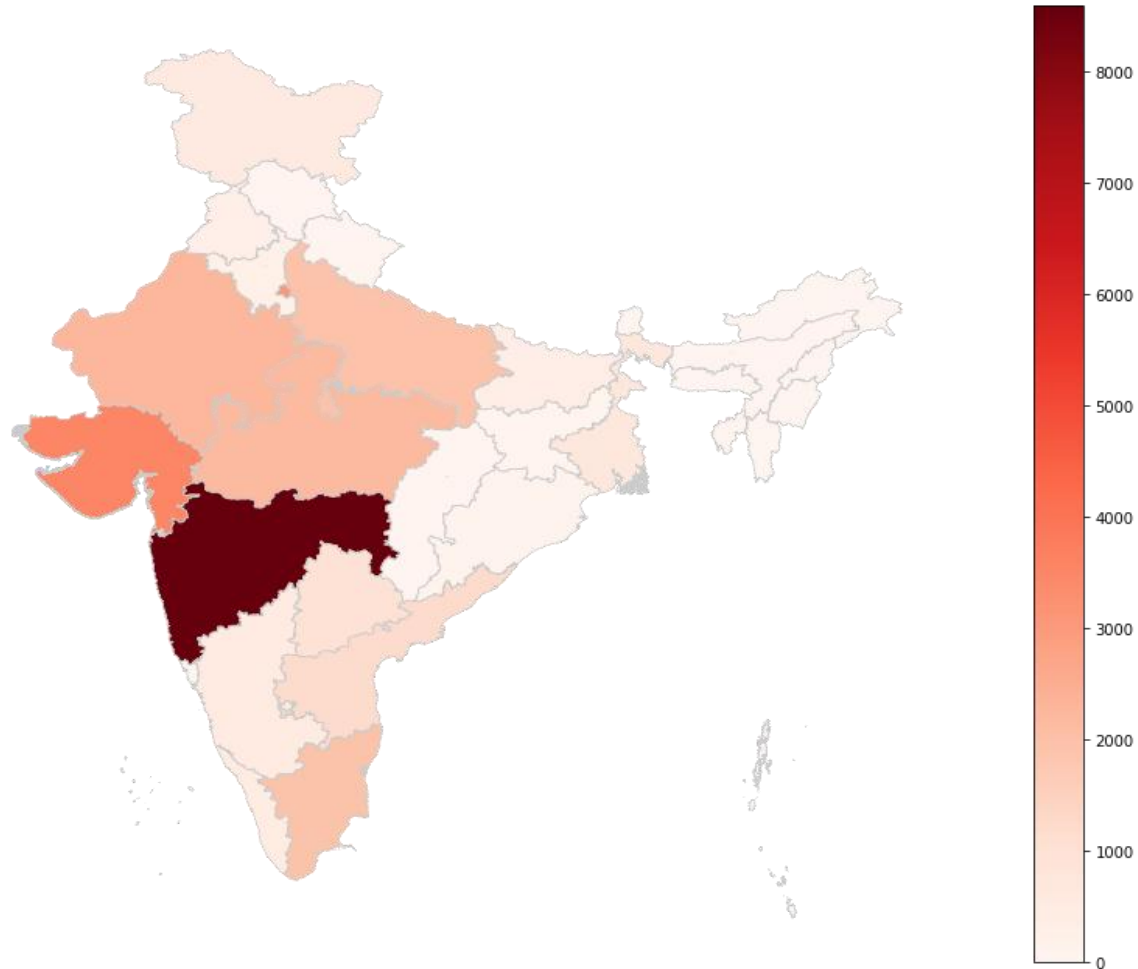


Figure: Proposed Solution Workflow

RUN-TIME Results

■ a. Indian States GeoTagged Covid-19 Hotspots Map

Covid-19 Statewise Statistics of Confirmed Cases



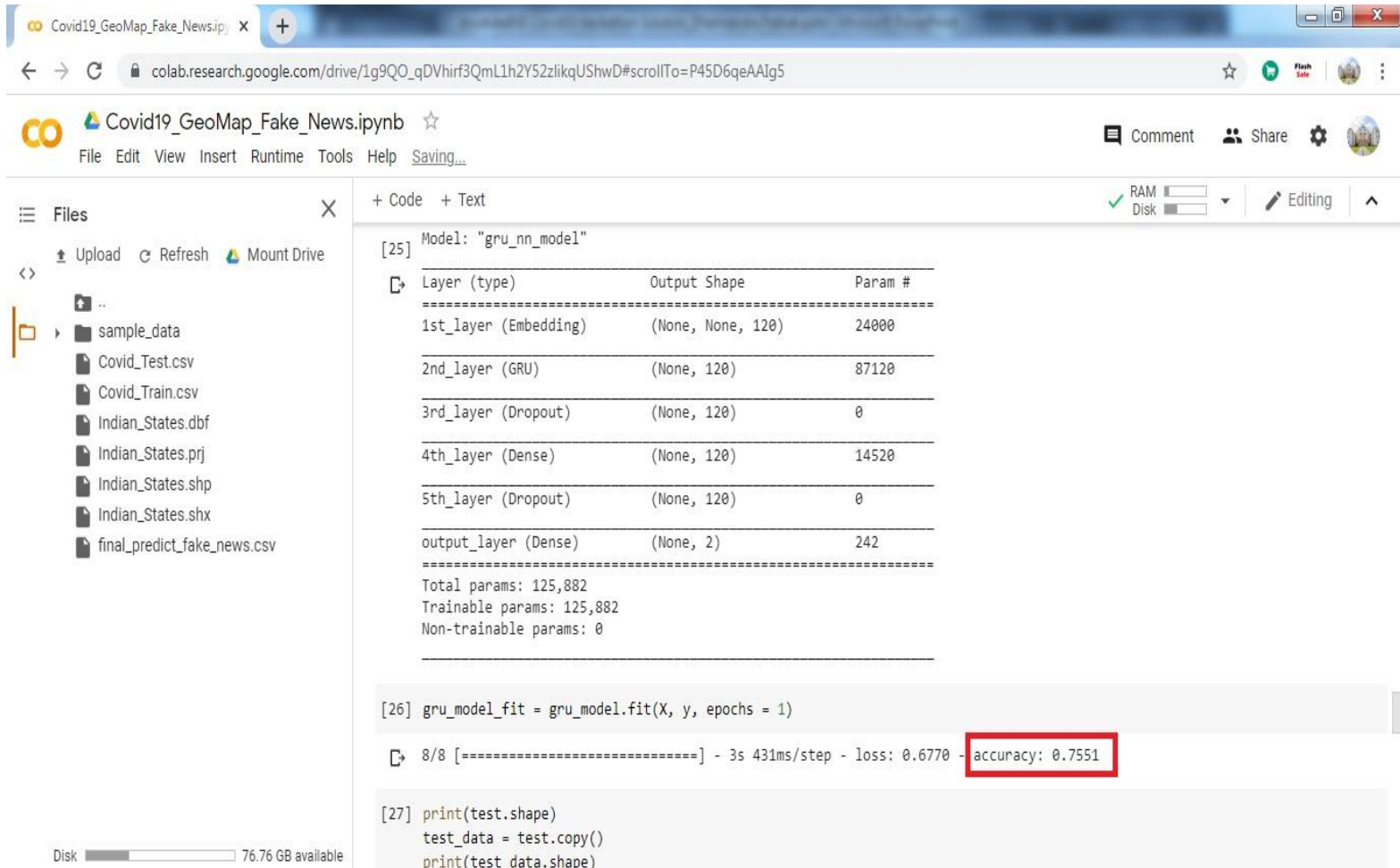
RUN-TIME Results

■ Covid-19 Twitter Data at a glance

1	created_at	id	text	user/id	user/name	user/location	user/url	lang	timestamp_ms	TARGET
2	Tue Apr 28 12:15:00	1.26E+18	The Bloomington City Council heard presentations from	19787029	WGLT	Normal, IL	http://wglit.org	en	1.59E+12	0
3	Tue Apr 28 12:15:00	1.26E+18	RT @gmlleunghku: The best exposition of the nuances	936139741	nicole king			en	1.59E+12	1
4	Tue Apr 28 12:15:00	1.26E+18	UK says some children have died from syndrome linked	44728980	ABS-CBN News	Philippines	http://news.abs-cbn.c	en	1.59E+12	1
5	Tue Apr 28 12:15:00	1.26E+18	#Amazon employees plan 'online walkout' to protest fir	15221598	James Gingerich	Waterloo, Ontario Can	http://www.linkedin.c	en	1.59E+12	1
6	Tue Apr 28 12:15:00	1.26E+18	Shielding your children from the #COVID19 crisis could	703192591	Chasing News with Bill Spadea		http://www.facebook	en	1.59E+12	0
7	Tue Apr 28 12:15:00	1.26E+18	The daily trends of #COVID19 cases in Bengaluru show	180748385	ORF	India	http://www.orfonline	en	1.59E+12	1
8	Tue Apr 28 12:14:56	1.26E+18	RT @MaajidNawaz: Dear pundits, I think this brief thre	1.23E+18	Wonko the Sane	Surrey		en	1.59E+12	0
9	Tue Apr 28 12:15:00	1.26E+18	Blackstone warns coronavirus could scupper a, ~1.36bn	525288897	Mike Didymus-T	London	http://www.altassets.	en	1.59E+12	1
10	Tue Apr 28 12:15:00	1.26E+18	@drharshvardhan Delhi government relaxes	231692806	Jagran English	New Delhi, India	http://english.jagran.c	en	1.59E+12	1
11	Tue Apr 28 12:14:58	1.26E+18	This is so dumb	95955741	((Daniel Ross))	Oklahoma, USA	http://www.christluth	en	1.59E+12	0
12	Tue Apr 28 12:15:00	1.26E+18	#COVID19 has disrupted access to #contraceptives, res	2255184786	Performance M	Baltimore, MD	http://www.pmdata.	en	1.59E+12	1
13	Tue Apr 28 12:15:00	1.26E+18	RT @BernardKerik: President @realDonaldTrump: The	9.74E+17	TasteTheBrainb	South Carolina, USA		en	1.59E+12	0
14	Tue Apr 28 12:15:00	1.26E+18	#CoronavirusIreland: 8VŠ— More rebates for motorists	72820731	Virgin Media Ne	Dublin, Ireland	https://www.faceboo	en	1.59E+12	1
15	Tue Apr 28 12:15:00	1.26E+18	RT @IndiaDST: Study shows that #COVID19 may affect	1.03E+18	Krish Kannan8V	Chennai, India		en	1.59E+12	1
16	Tue Apr 28 12:15:00	1.26E+18	Remember when #opioid abuse and #vaping were are	8.84E+17	We Insure Risk	New York, USA	http://weinsurerisk.cc	en	1.59E+12	1
17	Tue Apr 28 12:15:00	1.26E+18	PLEASE RETWEET	571878474	Doordarshan Na	New Delhi	http://doordarshan.gc	en	1.59E+12	1
18	Tue Apr 28 12:15:00	1.26E+18	RT @JAMA_current: Treatment guidelines for #ARDS st	768275844	vivere			en	1.59E+12	0
19	Tue Apr 28 12:15:00	1.26E+18	#COVID19 28th April Media Bulletin on Corona #StayHo	9.26E+17	Azeeta Seher Q	Jammu And Kashmir		en	1.59E+12	0
20	Tue Apr 28 12:15:00	1.26E+18	BOP Appeals Order to ID, Possibly Release 8VŠVulnera	314550535	The Business Jo	Youngstown, Ohio	http://businessjourna	en	1.59E+12	1
21	Tue Apr 28 12:15:00	1.26E+18	RT @NigelUNI: I was one of those who spoke to Panora	2851718342	Aine B			en	1.59E+12	0
22	Tue Apr 28 12:15:00	1.26E+18	RT @drbobgill: Wondering why UK will have 2nd	2724399753	YouCan'tBeSerious			en	1.59E+12	0
23	Tue Apr 28 12:15:00	1.26E+18	COVID-19: a guide to Australia8VŠ's main legal changes	145208608	LawInSport	United Kingdom	https://www.lawinspc	en	1.59E+12	1
24	Tue Apr 28 12:15:00	1.26E+18	RT @RachelNotley: This is so upsetting. @jkenney and	1.24E+18	Brenda Keating			en	1.59E+12	0
25	Tue Apr 28 12:15:00	1.26E+18	Stupid people elected this stupid person and now eve	9.24E+17	Rauhan 8VŠ+28VŠ+ MLE 8VŠ+28VŠ+»		https://www.faceboo	en	1.59E+12	0

RUN-TIME Results.....

- **b. Gated recurrent unit (GRU) deep learning algorithm Accuracy Snapshot for the training of Covid-19 Twitter Data with accuracy of 75.51 % for the detection of Fake News**



The screenshot shows a Google Colab notebook titled "Covid19_GeoMap_Fake_News.ipynb". The left sidebar shows a file explorer with a folder named "sample_data" containing several files: "Covid_Test.csv", "Covid_Train.csv", "Indian_States.dbf", "Indian_States.prj", "Indian_States.shp", "Indian_States.shx", and "final_predict_fake_news.csv". The main area displays the following code and output:

```
[25] Model: "gru_nn_model"
```

Layer (type)	Output Shape	Param #
1st_layer (Embedding)	(None, None, 120)	24000
2nd_layer (GRU)	(None, 120)	87120
3rd_layer (Dropout)	(None, 120)	0
4th_layer (Dense)	(None, 120)	14520
5th_layer (Dropout)	(None, 120)	0
output_layer (Dense)	(None, 2)	242

Total params: 125,882
Trainable params: 125,882
Non-trainable params: 0

```
[26] gru_model_fit = gru_model.fit(X, y, epochs = 1)
```

8/8 [=====] - 3s 431ms/step - loss: 0.6770 - accuracy: 0.7551

```
[27] print(test.shape)
test_data = test.copy()
print(test_data.shape)
```

The accuracy of 0.7551 is highlighted with a red box in the original image.



USP of Solution

- I. Real-time and synchronized tracking the impact of Covid-19
- II. Filtered data from the authentic source
- III. Better state-wise visualization through GeoTagging model
- IV. Trained Deep Learning model can be used for the identification of fake news and articles against large user twitter database
- V. Very helpful while taking further important decisions e.g. extending lockdown, food delivery, medical checkups etc.

Pricing, Time Line & Support Needed

- It is a machine learning based solution and implemented on open source APIs.
- The proposed idea has already been implemented.
- Hosted on Google Colab Platform, which is free of cost. Apart from this only requirement of the solution is High Configuration System (Processor+RAM+GPUs+SSD) for processing.

!! Thank you !!

