

Social Landslides



Jun 30, 2022

Badr Jaidi, Mariia Morello, Shengjie Zhang, Yiting Zhou

Contents

- **Project background and overview**
- **Intermediate tasks**
 - **Reddit article extraction**
 - **NASA database pre-processing**
 - **Location extraction**
 - **Date extraction**
- **Baseline and BERT multitask model**
- **Results**
- **Data product instructions**
- **Conclusion and future prospects**



Project Background



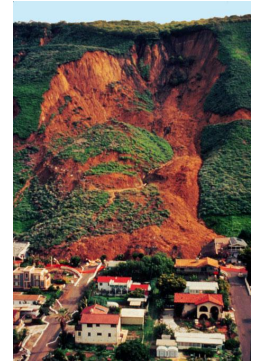
18 000 deaths
due to landslides

Between 1998-2017, landslides caused more than 18 000 deaths worldwide.

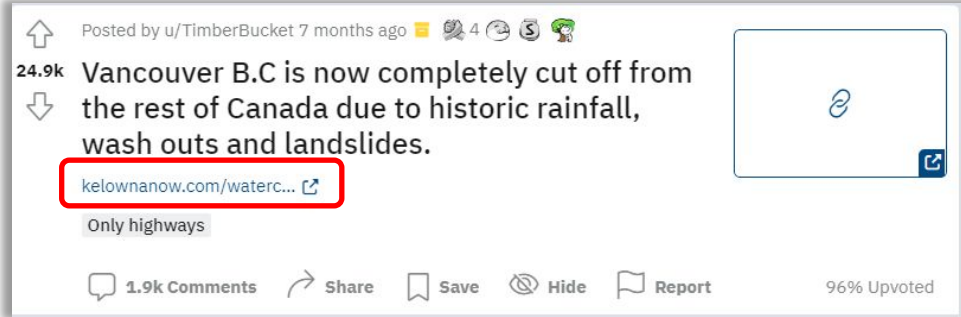


4.8 million people
affected

Between 1998-2017, landslides affected an estimated 4.8 million people worldwide.



Reddit Landslide-related Posts



Read URL links of news articles

Extract event information (location, date, time, trigger, death, injury numbers)

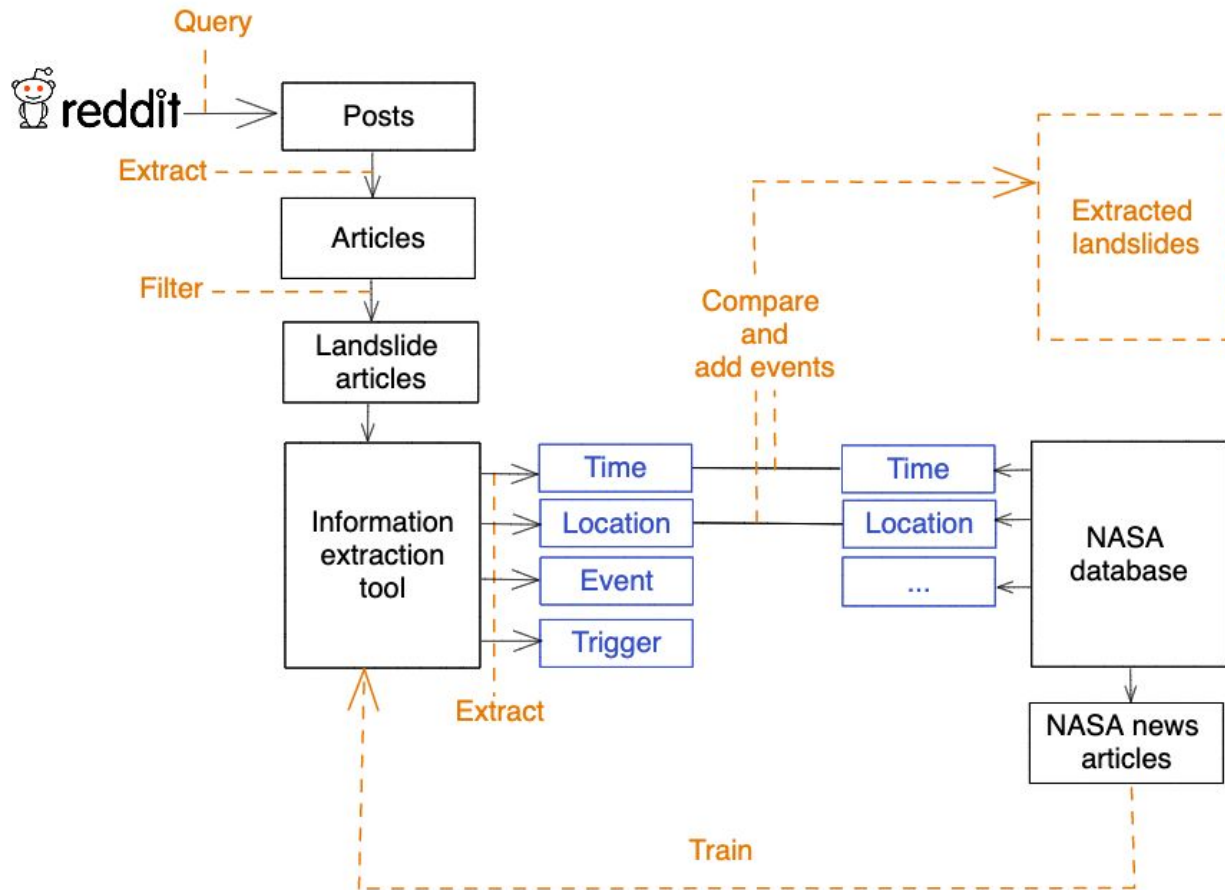
Reddit Landslide Event Information Detection



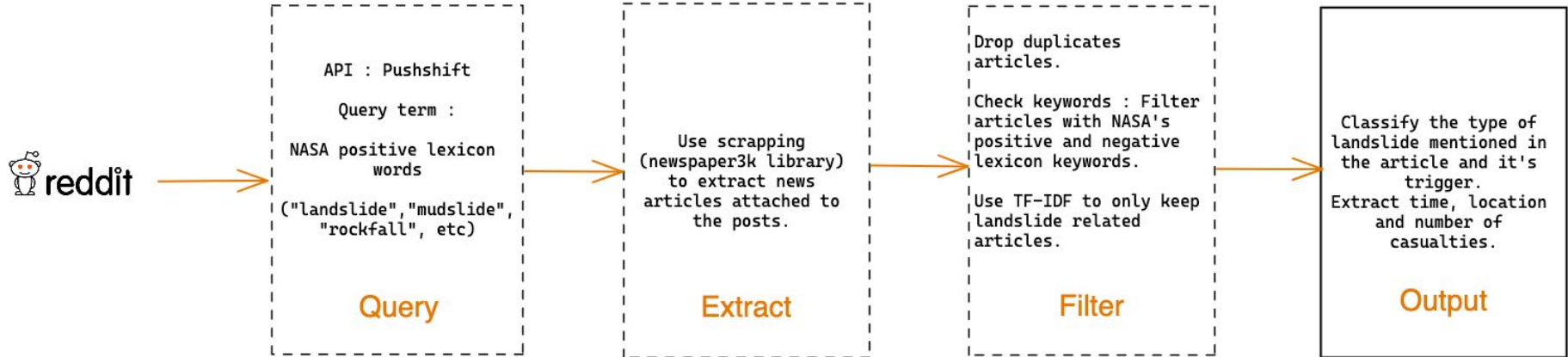
- **Location:** Barahbise
- **Date:** Sunday
- **Time:** before dawn
- **Death/injury count:** 12, 160
- **Landslide trigger:** torrential rain

KATHMANDU (Reuters) - Landslides triggered by **torrential rain** swept through two villages in **Nepal** on **Sunday**, **killing at least 12 people**, ... landslide swept away homes and people in **Barahbise**, 100 km (62 miles) east of Kathmandu near the border with the Tibet region of China, Nepali government official Murari Wasti said. Two others died in a landslide in Baglung in the northwest. Both landslides struck the villages **before dawn** ... At least 111 people remain missing and **160 have been injured**, Wasti said.

Project Overview



Reddit Article Extraction Pipeline



NASA Global Landslide Catalog Dataset



NASA's Open Data Portal

[Data Catalog](#) [About](#) [Developer Resources](#)



[Sign In](#)

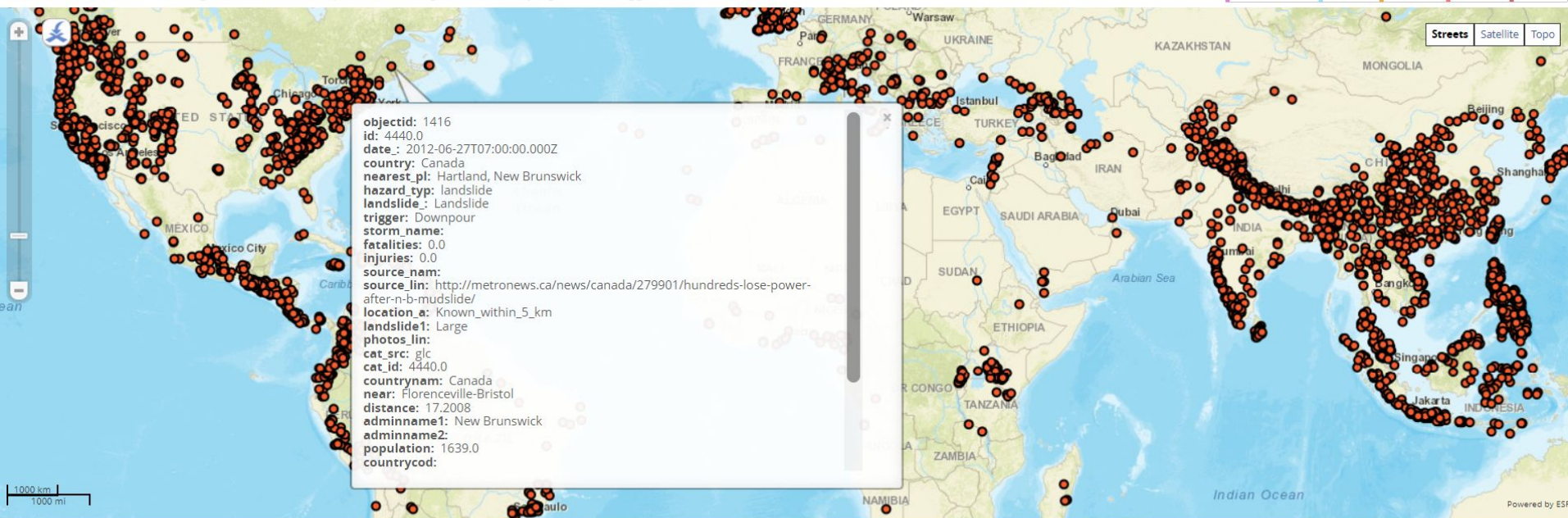
Global Landslide Catalog

Based on [Global Landslide Catalog](#)

The Global Landslide Catalog (GLC) was developed with the goal of identifying rainfall-triggered landslide



[More Views](#) [Export](#) [Discuss](#) [Embed](#) [About](#)



NASA Global Landslide Catalog Dataset



- Filtering by duplicate source links → 9000 examples
- Newspaper3k python library for scraping articles
- Filtering by newspaper articles → 5000 examples

Dataset Preprocessing



- Separating news texts into individual sentences
- Location and time information from each sentence by pre-train model

B	C	D	E	F	G	H
id	text	GPE	LOC	DATE	TIME	
0	Reported By: Edited By: Source: ANI Updated: Aug 04, 2015, 11:39 AM IST			4-Aug-15	11:39 AM IST	
0	At least 12 people are feared to be buried under the debris after three houses collapsed following a la	Dabhol Maharashtra Ratnagiri				
0	The landslide is said to have taken place at 3:30 am due to heavy rainfall.				3:30 AM	
0	Ratnagiri (Maharashtra): Landslide occurred at around 3: 30 am today.	Ratnagiri Maharashtra			3 : 30 am today	
0	pic.twitter.com/MOcmTxnAZX June 22, 2015			22-Jun-15		
0	Spot visuals: Landslide in Dabhol Village, Ratnagiri (Maharashtra) due to heavy rainfall; rescue operati	Dabhol Village Ratnagiri Maharashtra				
0	pic.twitter.com/a5mMtl2hwB ,Â@ANI (@ANI_news) June 22, 2015			22-Jun-15	Earlier in the day	
1	Gilgit/Hunza/Skardu: At least eight people were killed and around two dozen were injured across Gilgi	Gilgit-Baltistan		today		
1	Having its epicenter in Afghanistan, the earthquake shook the entire mountainous GB region, spreadin	Afghanistan GB				
1	[dropshadowbox align=,Âuright,Âu effect=,Âuhorizontal-curve-bottom,Âu width=,Âu200px,Âu height=,Âu	Gilgit-Baltistan				
1	Diamer (13), Ghizar (09), Gilgit (04), Hunza (02), Nagar (01), Ghanche (03) ,Âi Abdul Rehman	Gilgit Hunza Ghizar Distri (09) 04)				
1	One source has put the death toll in Ghizar at 4.	Ghizar				
1	The deaths occurred when a vehicle was reportedly hit by a boulder while traveling on the Gilgit-Shanc	Gilgit-Shandur				
1	Meanwhile, there are also reports that two children were severely injured in Thoi village of Yasin Tehs	Thoi village Yasin Tehsil				
1	A car carrying five people of a family was hit by a landslide or rockfall in the Kohistan area.	Kohistan				
1	Three of the passengers are reportedly missing, while dead body of one has been recovered from the Indus river.	the Indus river				
1	The people traveling in the car were residents of Danyore Town of Gilgit district.	Danyore Town Gilgit				
1	In Diamer district two people, including a child, were killed while at least 13 people sustained minor a	Diamer Darel Tangir Tatta Pani				
1	In Hunza, a female teacher died after being hit by a landslide between Ganish and Sumayar.	Hunza Ganish Sumayar				
1	She was returning home from a school in Sumayar (Nagar), where she taught.	Sumayar Nagar				
1	The landslide threw white dust across the valley, while also blocking the road to Nagar Khas.	Nagar Khas				
1	However, the landslide debris did not affect flow of the Hunza River.	the Hunza River				
1	Boundary walls fell and some buildings got cracks in Gilgit as a result of the tremors.	Gilgit				
1	Main and link roads have been blocked by landslide and rockfall across the GB region	Astore District				
1	A couple of people were injured in Ghanche district	Ghanche				

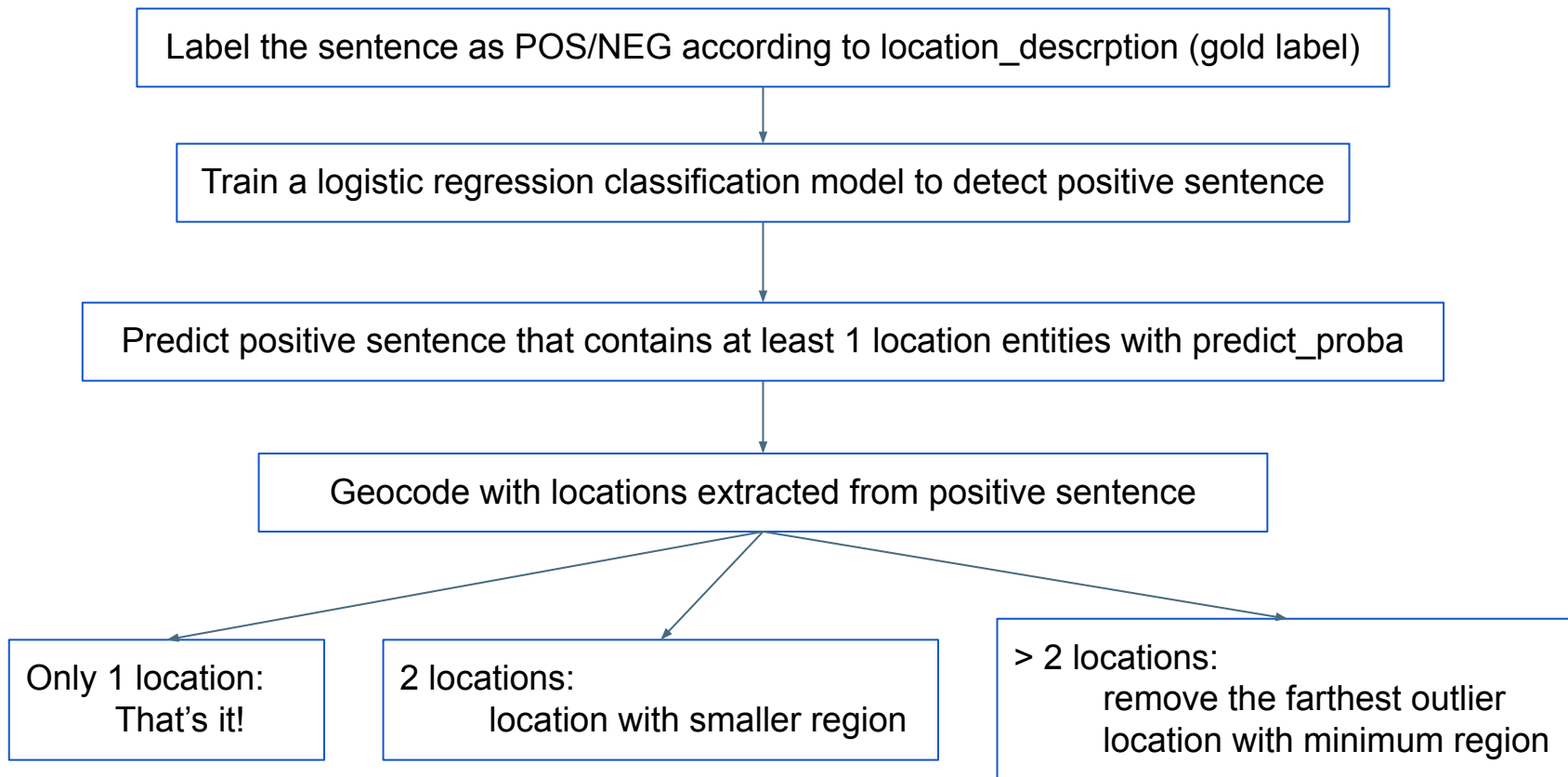
Location Extraction: Model



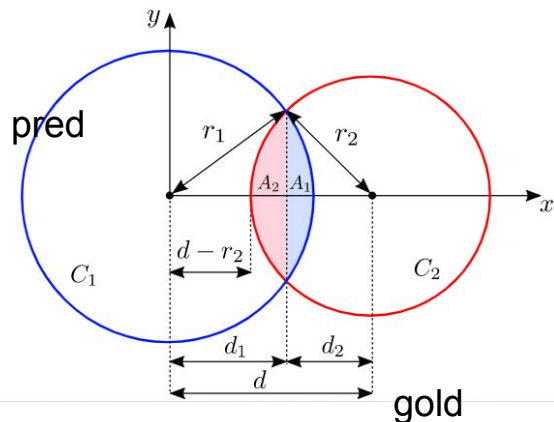
KATHMANDU (Reuters) - Landslides triggered by torrential rain swept through two villages in **Nepal** on Sunday, killing at least 12 people, ... landslide swept away homes and people in **Barahbise**, 100 km (62 miles) east of **Kathmandu** near the border with the **Tibet** region of **China**, **Nepali** government official Murari Wasti said. Two others died in a landslide in **Baglung** in the northwest. Both landslides struck the villages before dawn ... At least 111 people remain missing and 160 have been injured, Wasti said.

- **Extract all the locations from the text**
 - use a NER model!
- **Find the right one with a model**
 - built a logistic regression model to get the sentence containing the right location
- **Transform the location into latitude/longitude**
 - ArcGIS geocoder tool

Location Extraction: Model



Location Extraction: Evaluation Metrics



Precision = intersection_area / predicted_area
 Recall = intersection_area / gold_area
 F1_score = 2*Precision*Recall/(Precision+Recall)

Summary

Given two circles C_1 and C_2 of radii r_1 and r_2 respectively (with $r_1 \geq r_2$) whose center points are at a distance d from each other, the intersection area of the circles is:

1. zero, if $d \geq r_1 + r_2$, since in this case the circles intersect at most up to a point.
2. πr_2^2 , if $d \leq r_1 - r_2$, since in this case C_2 is entirely contained within C_1 .
3. given by equation (14) in all other cases.

The sum of A_1 and A_2 is the intersection area of the circles:

$$A_{\text{intersection}} = r_1^2 \cos^{-1}\left(\frac{d_1}{r_1}\right) - d_1 \sqrt{r_1^2 - d_1^2} + r_2^2 \cos^{-1}\left(\frac{d_2}{r_2}\right) - d_2 \sqrt{r_2^2 - d_2^2} \quad (14)$$

where:

$$d_1 = \frac{r_1^2 - r_2^2 + d^2}{2d} \quad \text{and} \quad d_2 = d - d_1 = \frac{r_2^2 - r_1^2 + d^2}{2d} \quad (15)$$

Location Extraction: Evaluation Metrics



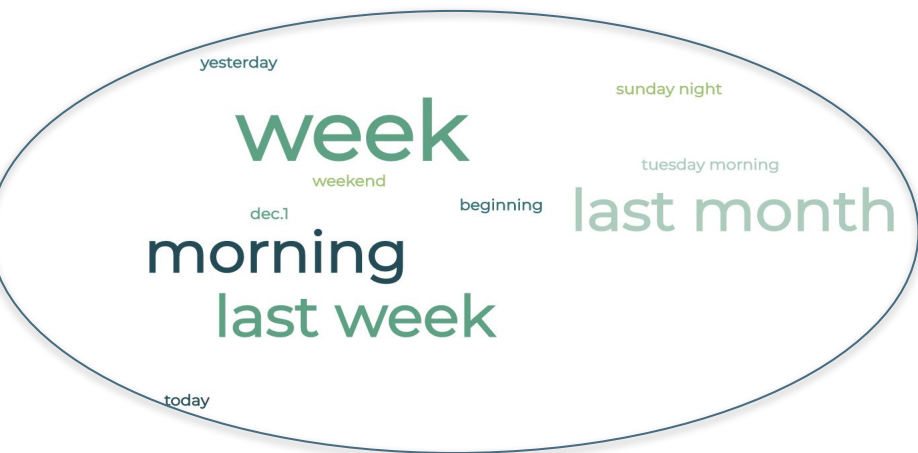
id	locations	location	latitude	longitude	radius_km	location_description	location_accuracy	gold_latitude	gold_longitude	distance_km	gold_radius_km	correct	precision	recall	f1_score
27116	Purul-Lakhamai Senapati Paokho	Senapati	25.26867	94.02224	3.1445	near the Khongjaron Khunthak village, Tamei Ro...	10km	24.938948	93.496837	64.362	10.0	False	0.000000	0.000000	0.000000
27117	Kohima	Kohima	25.66288	94.09311	6.8775	P.R. Hill, Kohima, Nagaland 797005, India.	1km	25.658708	94.098536	0.714	1.0	True	0.021142	1.000000	0.041408
27118	Kot village Tehri	Kot village	29.52405	81.58881	1.4725	SH-8, Tehri Sub-District, Tehri Garhwal, Uttar...	1km	30.359485	78.377428	323.556	1.0	False	0.000000	0.000000	0.000000
27132	West Java Indonesia	West Java	-6.75000	107.50000	171.0650	Cipelah, Rancabali, Bandung, West Java, Indone...	5km	-7.190355	107.289860	53.952	5.0	True	0.000854	1.000000	0.001707
27133	Princeton	Princeton	40.34954	-74.66009	1.8175	Mountainside, off of Highway 5A, 8 miles from ...	1km	49.526485	-120.521235	3698.089	1.0	False	0.000000	0.000000	0.000000
...
35610	Mishawaka	Mishawaka	41.66017	-86.17217	7.9130	On County Road 27 -- The Stove Prairie Road --...	5km	40.655800	-105.374200	1612.212	5.0	False	0.000000	0.000000	0.000000
35613	Tamenglong Daniel Kamei Tamenglong	Tamenglong	24.96538	93.49509	23.8325	Tamenglong Ward No-IV	50km	24.981800	93.514200	2.652	50.0	True	1.000000	0.227195	0.370267
35633	Surigao del Sur DAVAO CITY Philippines Surigao...	DAVAO CITY	7.06574	125.61080	46.7400	Balibadon, Surigao del Sur	5km	9.234000	126.169600	247.582	5.0	False	0.000000	0.000000	0.000000
35646	PAANCHTHAR Phidim-Raanke Samdin Phidim-2	Phidim-Raanke	27.12615	87.75943	1.4870	Phidim-Raanke section of the Mechi Highway at ...	10km	27.100500	87.783100	3.686	10.0	True	1.000000	0.022112	0.043267
35649	Jammu	Jammu	32.70273	74.87870	14.3775	Chakwa bridge, 9 kilomters short of Doda	5km	33.126100	75.448100	71.004	5.0	False	0.000000	0.000000	0.000000

555 rows × 15 columns

Date/time extraction and normalization



id	text	GPE	LOC	DATE	TIME
0	Reported By: I Edited ...			Aug 04 , 2015	11:39 AM IST
0	At least 12 people are...	DabholMaharashtral...			
0	The landslide is said t...				3:30 am
0	Ratnagiri (Maharashtr...	RatnagiriMaharashtra			3 : 30 am today
0	pic.twitter.com/MOcm...			June 22 , 2015	
0	Spot visuals: Landslid...	Dabhol VillagelRatna...			
0	pic.twitter.com/a5mMt...			June 22 , 2015	Earlier in the day



Date/time extraction and normalization



Year-phrases

'in 2012'
'last year'
'two years ago'

Season-phrases

'last autumn'
'summer'
'last summer'

Month-phrases

'last month'
'this month'
'last August'
'beginning of July'
'middle of February'
'late September'

Week-phrases

'last week'
'this week'
'first week of March'
'second week of May'
'last week of April'
'beginning/middle/end of last week'

Weekend-phrases

'last weekend'
'weekend'
'this past weekend'
'middle of last weekend'
'early past weekend'

Parts-of-day-phrases

'last morning'
'this evening'
'early in the afternoon'
'late last night'

Time-phrases

'4 AM'
'between 12 am and 5 p.m.'
'around 5:15 a.m.'

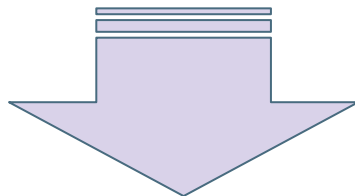
Date/time extraction and normalization



'last August' - which day and what time?

'last weekend' - what time?

'in 2017' - what month, what day and what time?



'last August' - (12AM 1st of August - 23:59PM 31st of August, 2021)

'last weekend' - (12AM 25th of June - 23:59PM 26th of June, 2022)

'in 2017' - (12AM 1st of January - 23:59PM 31st of December, 2017)

Date/time extraction and normalization



Publication date - anchor

Gold date/event date - the NASA GLC event date

Ex. 23rd of March, 2022(Wednesday) - 23/03/2022 00:00

'yesterday' => 22nd of March (22/03/2022 0:00 - 22/03/2022 23:59)

'morning' => 4AM - 11AM 22nd of March (**22**/03/2022 4:00 - **22**/03/2022 11:00)

'Tuesday morning' => 4AM - 11AM 21st of March (21/03/2022 4:00 - 21/03/2022 11:00)

'last month' => February (01/02/2022 00:00 - 28/02/2022 23:59)

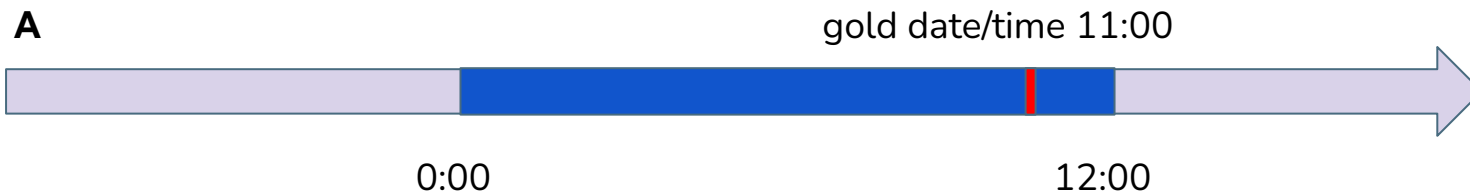
Date/time extraction and normalization



Evaluation - binary accuracy

Gold date 23/03/2022 11:00

Normalized date interval(A) - 23/03/2022 00:00 - 23/03/2022 12:00



Normalized date interval(B) - 23/03/2022 12:00 - 23/03/2022 18:00

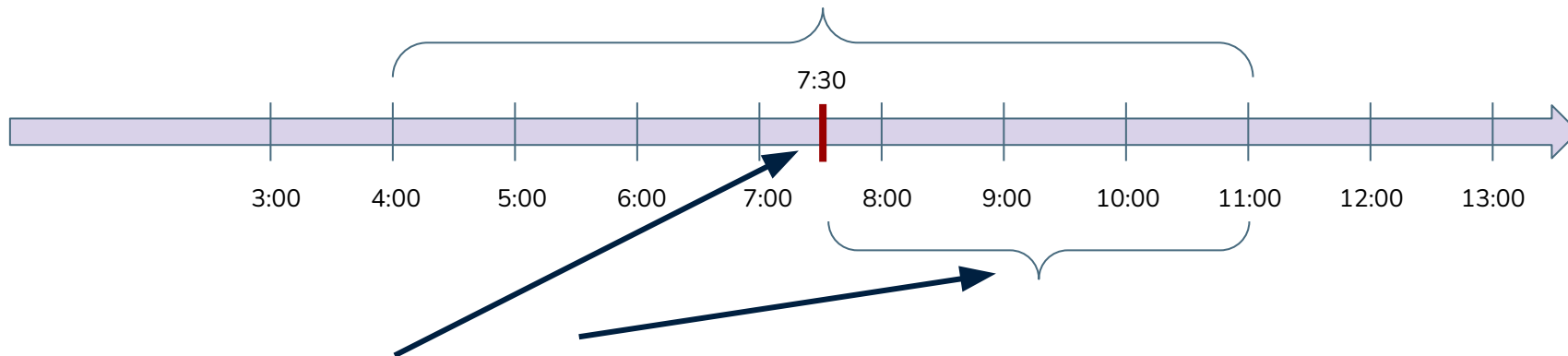


Date/time extraction and normalization



Initial goals: discrete date + confidence level

Final output: given an interval - 28/03/2022 4:00 - 28/03/2022 11:00



Return discrete date + confidence:

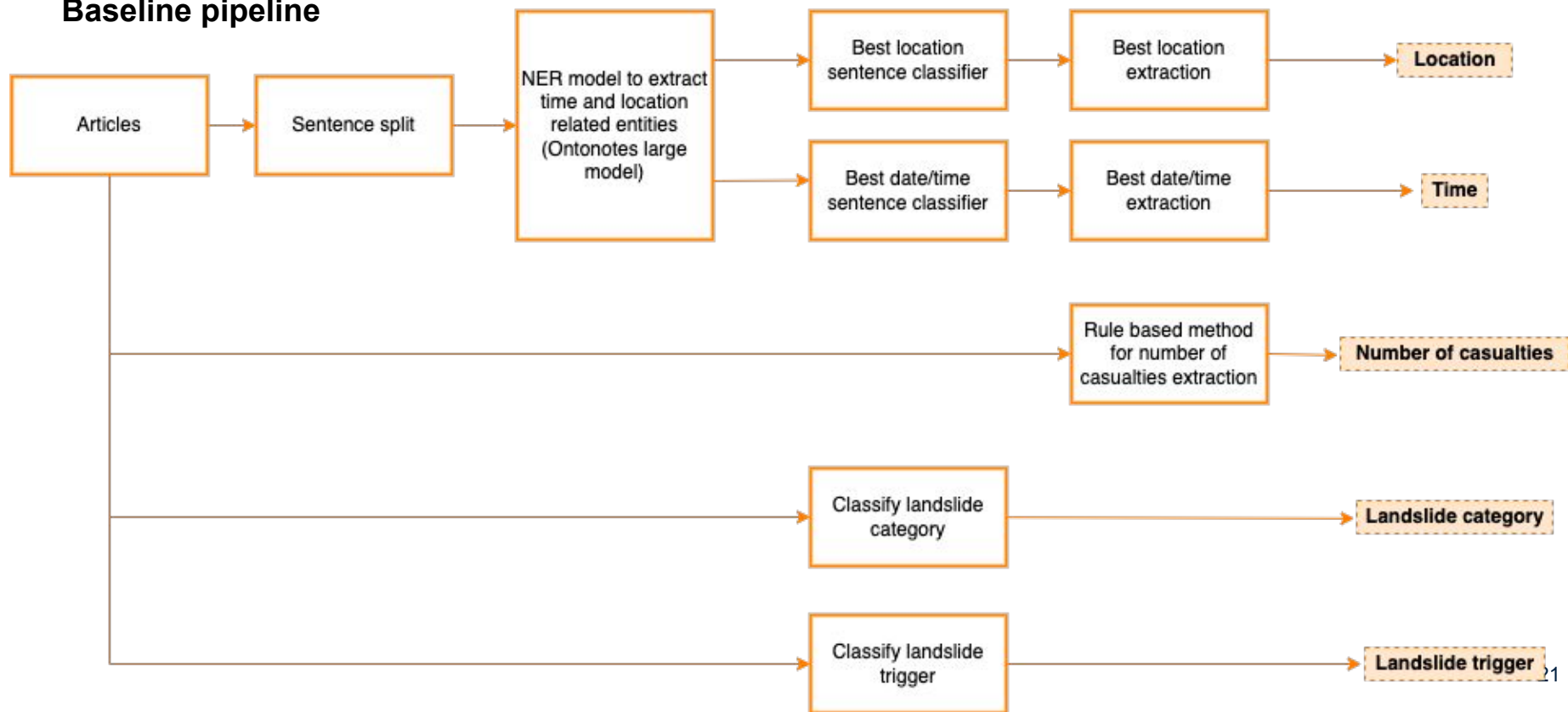
28/03/2022 7:30 \pm 3.5 hours

Landslide Event Extraction



Logistic regression classification

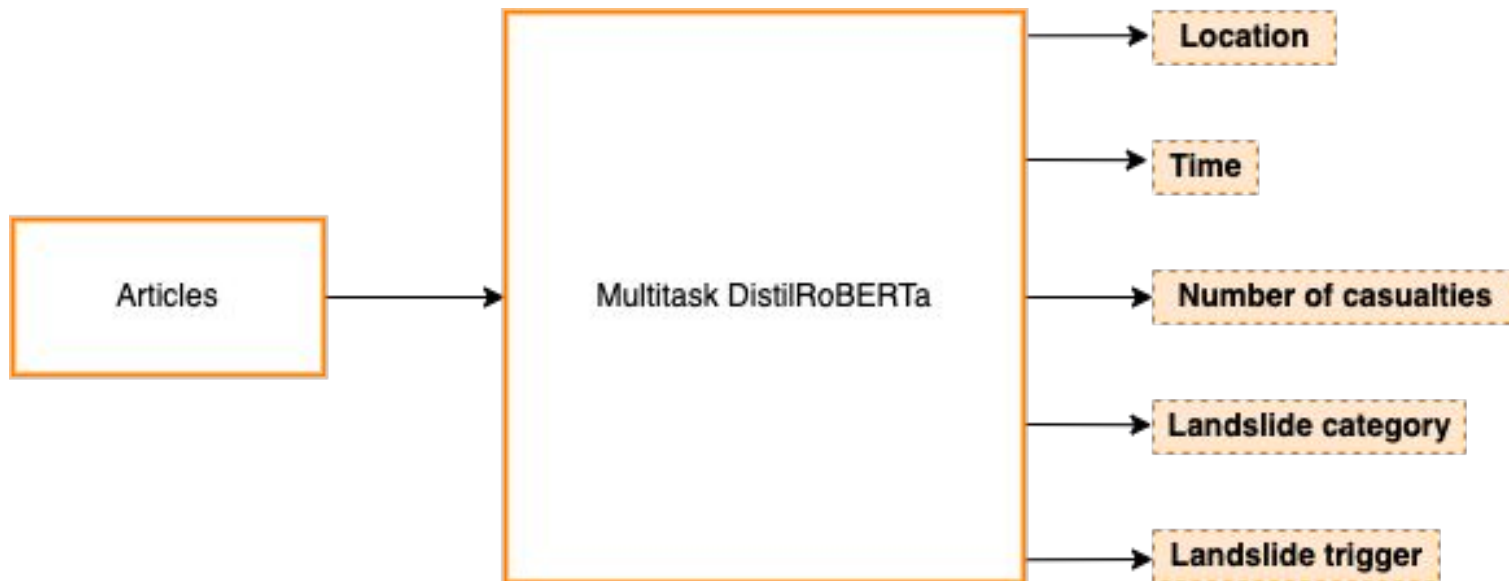
Baseline pipeline



Landslide Event Extraction



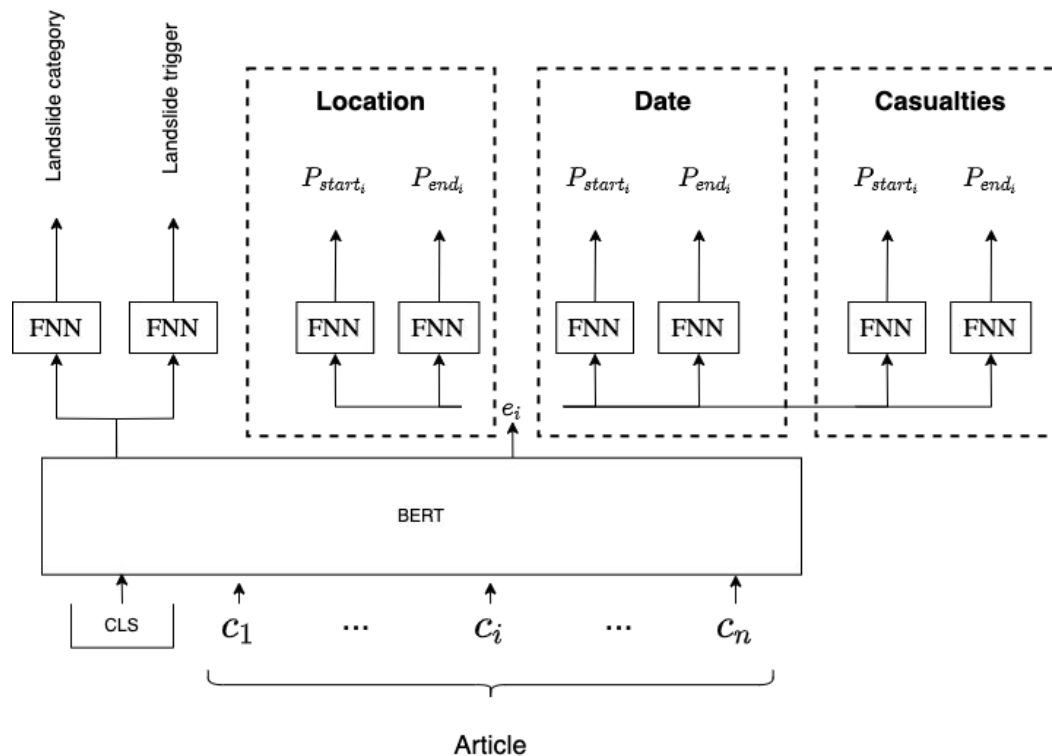
Multitask pipeline



Landslide Event Extraction



Multitask model



Results



Baseline advantages:

- Much faster to train.
- Chooses the best time and location out of multiple candidates.

	LR		MT	
	dev	test	dev	test
category (f1)	0.7484	0.7174	0.8037	0.7950
trigger (f1)	0.4534	0.4503	0.4579	0.4480
casualties (f1)	0.4030	0.4160	0.4309	0.4528
location (acc)	0.4708	0.4452	0.4579	0.4480
date/time (acc)	0.3428	0.4224	0.3568	0.3418

Multitask advantages:

- Much faster to run.
- Simpler implementation, everything is extracted directly.

Extracting events from 40 articles:

Baseline : 4 mins

Multitask : 40 seconds

Examples



- **Location:** Boragaon
- **Date:** Wednesday
- **Death/injury count:** 4
- **Landslide category :** mudslide
- **Landslide trigger:** continuous rain

Fresh landslides were reported from various parts of Guwahati on Wednesday, amid incessant rainfall, which also led to heavy flooding in most parts of the city, officials said.

During the day there were no reports of new casualties.

The state has recorded 42 deaths this year from flooding and landslides, including **four** that were buried under a massive mound of earth in the Boragaon area on Tuesday.

Examples



- **Location:** Jiangxi
- **Date:** this week
- **Death/injury count:** 6
- **Landslide category :** landslide
- **Landslide trigger:** downpour

BEIJING -- Heavy rain in China has claimed **six** lives **this week** and forced the evacuation of some 200,000 people.

Five people died in southeastern China, according to state media reports: A house collapse took three lives in Fujian province on Tuesday, and two died Monday in a landslide caused by torrential rain in neighboring **Jiangxi** province.

Examples



- Location: Yellowstone National Park
- Date: Monday morning
- Death/injury count: 0
- Landslide category : debris flow
- Landslide trigger: rain

YELLOWSTONE NATIONAL PARK - Roads in the northern portion of **Yellowstone National Park** were temporarily closed **Monday morning** due to substantial flooding, rockslides and mudslides on roadways from recent unprecedented amounts of rainfall.

Examples



- **Location: Wyoming**
- **Date: Sunday**
- **Death/injury count: 0**
- **Landslide category : rockslide**
- **Landslide trigger: downpour**

Visitors to Yellowstone National Park were seen narrowly escaping with their lives after a rockslide began falling on their vehicle amid devastating floods.

The drama took place when heavy rain fell on northern **Wyoming** and southern Montana on **Sunday**, causing record-breaking flooding into Monday and Tuesday.

Final Product Pipeline



- Scraping Reddit data
- Predicting landslide events
- Verifying new events and deleting duplicate events

Final Product



- Docker image tar file
- Config file

```
{
  "interval" : {
    "default" : "no",
    "start" : {
      "date" : "2022-04-01"
    },
    "end" : {
      "now" : "no",
      "date" : "2022-04-21"
    }
  },
  "model" : {
    "baseline" : "yes",
    "multitask" : "yes"
  }
}
```

Installing instruction



- Install docker
- Download config file and main product tar file into local file system
- Change the content of the config file so user can customize the time span and the model type
- Go to directory where you save the main file in terminal
- Type ‘docker load < landslides.tar’
- Type ‘docker run -v \$(pwd)/user:/user landslides’
- After running the application, the results should save in files folder



config.json



history.log



results_bert.csv

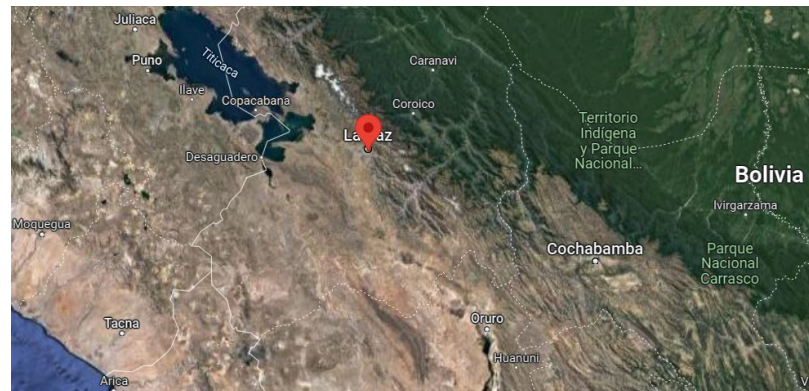


results.csv

Landslide Event Information Detection and Extraction Tool



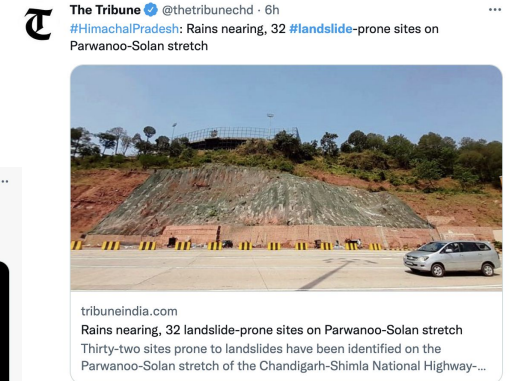
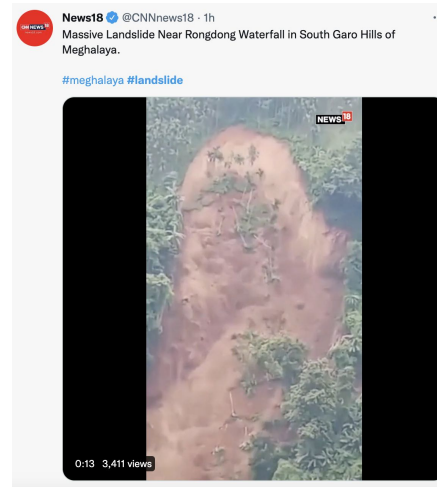
- 1) Location
- 2) Location accuracy/confidence
- 3) Date/time
- 4) Date/time accuracy
- 5) Number of casualties
- 6) Trigger
- 7) Category
- 8) Duplicates in the database



What is next?



- expand to other social media platforms
- text data: real time and users posts
- other types of data: images and videos
- more complex systems to boost the accuracy





THE UNIVERSITY OF BRITISH COLUMBIA