

Extracting and Exploring Information about Flood Events from Twitter

Research Questions

- How to classify relevant tweets?
- How to extract locations from tweets?
- What insights can be extracted from tweets' text?
- What visualizations can be used to represent the results?

Disaster management



Figure: Floods in Pakistan 2022¹

¹https://en.wikipedia.org/wiki/2022_Pakistan_floods

Disaster management

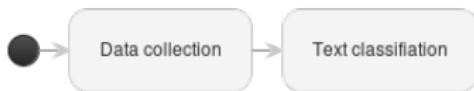


How can it facilitate disaster management?

- Preparation
- Response
- Recovery

Figure: Floods in Pakistan 2022¹

¹https://en.wikipedia.org/wiki/2022_Pakistan_floods



■ Unlabelled data:

- Extracted from Twitter API using flood relevant terms
- Analyse historical flood events

■ Labelled data:

- Crowdsourced datasets (25000 tweets)
- Train DistilBERT to classify relevant tweets

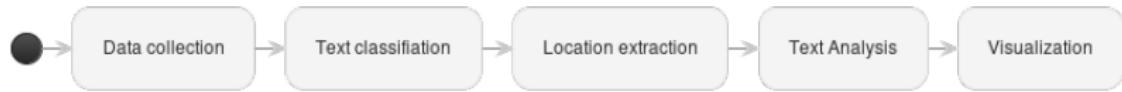


- Geoparsing process:

- **Toponym recognition** using an NER model
- **Toponym resolution** using OpenStreetMap
- Given two or more locations in one tweet, select the location with the smallest bounding box

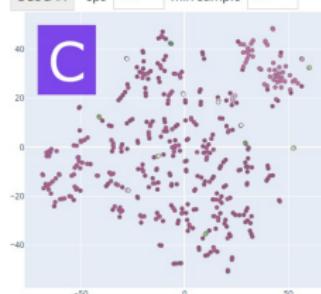
- Text analysis techniques:

- Topic modelling using **LDA**
- Words relevance using **TF-IDF**
- Dimensionality reduction using **t-SNE** on TF-IDF matrix with DBSCAN clustering



<input type="checkbox"/> id	<input checked="" type="checkbox"/> raw	<input checked="" type="checkbox"/> translated	<input checked="" type="checkbox"/> processed	<input type="checkbox"/> hashtags	<input type="checkbox"/> loc_name	<input type="checkbox"/> softmax	<input type="checkbox"/> created_at
processed	raw	translated					
flood location himmelstalund norrköping municipality	Oversvämning: Plats: Himmelstalund - Norrköpings kommun https://t.co/AO3SE9h1Np	Flood: Location: Himmelstalund - Norrköpings municipality http://t.co/AD3SE9h1Np					
flooding marihamn	Oversvämning i Marihamn https://t.co/8Q0MhB6ic http://t.co/gtF494p2U	Flooding in Marihamn http://t.co/RDQ0B6M8ic http://t.co/gtF494p2U					
yes forgot polar ice caps conflict russia famine storms sea floods isis plastic sea greece	Ågårdsposten Ja. Vi glömmer polarisarna, konflikten med Ryssland, svält, stormar från havet, översväningar, Isis, plasti havet, Grekland..	Ågårdsposten Yes. We forgot the polar ice caps, the conflict with Russia, famine, storms from sea, floods, Isis, plastic in the sea, etc..					
jämtland västerorland gävleborg 1km weather smhi warning class high flows small...	Jämtland, Västerorland och Gävleborg län (Väder): SMHI varnин klass 1 hög flöden i små v. http://t.co/PzglnH7Eur	Göteborg (Väder): SMHI warning class 1 high flows 1 and vattendräg i När flöden i små v. http://t.co/8y30uktOpG					
gothenburg weather smhi warning class high flows small waterways ditches information	Göteborg (Väder): SMHI warning klass 1 höga flöden 1 och vattendräg i När flöden i små v. http://t.co/8y30uktOpG	Göteborg (Weather): SMHI warning class 1 high flows in small waterways and ditches. Information http://t.co/8y30uktOpG #Öster #trafikP4					
e kungsbackaleden traffic disruption flooding disturbance olskrokmetot and..	E6 Kungsbackaleden (Trafikstörning): Översvämning. Störning Mellan Olskrokmetot (74) och.. http://t.co/R4VWVMOR1 #väg #trafikP4	E6 Kungsbackaleden (Traffic disruption): Flooding. Disturbance between Olskrokmetot (74) and.. http://t.co/R4VWVMOR1 #väg #trafikP4					
	Alltså.. Gårdagen. För mkt vin,	So.. Yesterday. Too much wine, flooding in the					

<< < > >>



Text analysis # topics: 2

LDA Topics TF-IDF terms

#term	#count	#topic0	#topic1
flooding	111	0.00029	1.00085
rain	67	0.0007	1.01426
traffic	65	0.00173	0.99983
high	52	1.0389	0
flow	51	0.99962	0
close	50	0	1.01998
warn	49	0.99718	0
smhi	48	0.99936	0
road	47	0.00082	1.02103
disruption	47	0.01246	1.21234
lets...	47	0.00048	0



Tweets: Total: 1190, With location: 423, Selected: 423 ,Has word Sweden: 34 ,
Spans: from 2015-07-08 12:14:12 to 2021-09-25 13:00:54, Locations: Total: 229, Selected: 229 ,
Kina (16), Tyskland (13), Gävle (13), Gävleborg County (11), Skåne County (9), etc.



Figure: Visual interface

Experiment to validate the pipeline

Classifier's accuracy (0.92), precision (0.9), recall (0.92) - trained on 20,000 tweets.

Extract one week's worth of tweets about a past flood event for a past event in Gävleborg and Dalarna counties on 18 August 2021

- **Misclassified Tweet:** A basement was flooded when a water main leaked in #Vårberga in #Borgå #borgåvatten
- **Misidentified location:** Dödssiffran stiger i turkiska översvämnningar
#**Turkiet** #svpol <https://t.co/K6kLRmxQdw>
Identified location: Turkiet, a hamlet² in Uppsala county.
Actual location: Turkey, the country.

²isolated settlement

Short Demo

Results Interpretation, Limitation and Future Word

- Add more data sources
 - Other social media platforms
 - Media outlets
 - Meteorological data
- Classifier performance
 - Process other elements, such as Images and URLs
- Identifying geographical locations
 - Heuristics to identify correct location
- Future Work
 - Forecasting
 - Other types of disasters
 - Other countries

Summary and Last Words

- The pipeline is able to extract information about historical flood events
- Social media is a potential data source to augment disaster management pipelines but not as a standalone source
- Highly dependent on people's participation
- Potential framework acknowledged by the people to motivate them to share their knowledge