Extracting and Exploring Information about Flood Events from Twitter

Yasser Kaddoura

Department of Information Technology
Uppsala University

Supervisors:
Carlo Navarra[1], Katerina Vrotsou[2], Kostiantyn Kucher[2]
[1,2]Linköping University, [1]Department of Thematic Studies,
[2]Department of Science and Technology

Thesis Defence
April 20, 2023

UPPSALA
UNIVERSITET

Contents

UPPSALA
UNIVERSITET

Flood Events

Figure: Floods in Dalarna and Gävleborg 2021[2]

Figure: Floods in Pakistan 2022[3]

[2] https://floodlist.com/europe/central-sweden-floods-august-2021
[3] https://en.wikipedia.org/wiki/2022_Pakistan_floods

Motivation and Research Questions

How can it facilitate disaster management?

- Preparation
- Response
- Recovery

Motivation and Research Questions

How can it facilitate disaster management?

- Preparation
- Response
- Recovery

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?

Data Sources

- **Labelled data**:
  - Crowdsourced datasets with “on-topic” and “off-topic” labels (25000 tweets)
  - Train the classifier

Data Sources

- **Labelled data**:
  - Crowdsourced datasets with "on-topic" and "off-topic" labels (25000 tweets)
  - Train the classifier
- **Unlabelled data**:
  - Extracted from Twitter API using the start time, end time, and a query
  - Analyse historical flood events

Portion of the query used

"hög vatten" OR åskskur OR vattennivå OR åskväder OR regnstorm OR "mycket regn" OR översvämningsskador OR översvämningar OR översvämning

English translation

"high water" OR thunderstorms OR "water level" OR thunderstorms OR rainstorm OR "a lot of rain" OR "flood damage" OR floods OR flood

Text Classification

- The terms used to fetch the tweets can have other "flood" can be used figuratively (e.g. flood of joy)

Text Classification

- The terms used to fetch the tweets can have other "flood" can be used figuratively (e.g. flood of joy)
- Transformers
  - Transfer Learning
  - Encoder-decoder architecture
  - Self-attention mechanism
- DistilBERT, a variant of BERT

Location Extraction

- Only 1% of tweets are geotagged by users

Location Extraction

- Only 1% of tweets are geotagged by users
- Geoparsing process:
  - **Toponym recognition** using an NER model
  - **Toponym resolution** using OpenStreepMap
- Given two or more locations in one tweet, select the location with the smallest bounding box

Text Analysis

Text analysis techniques:

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

Visualization

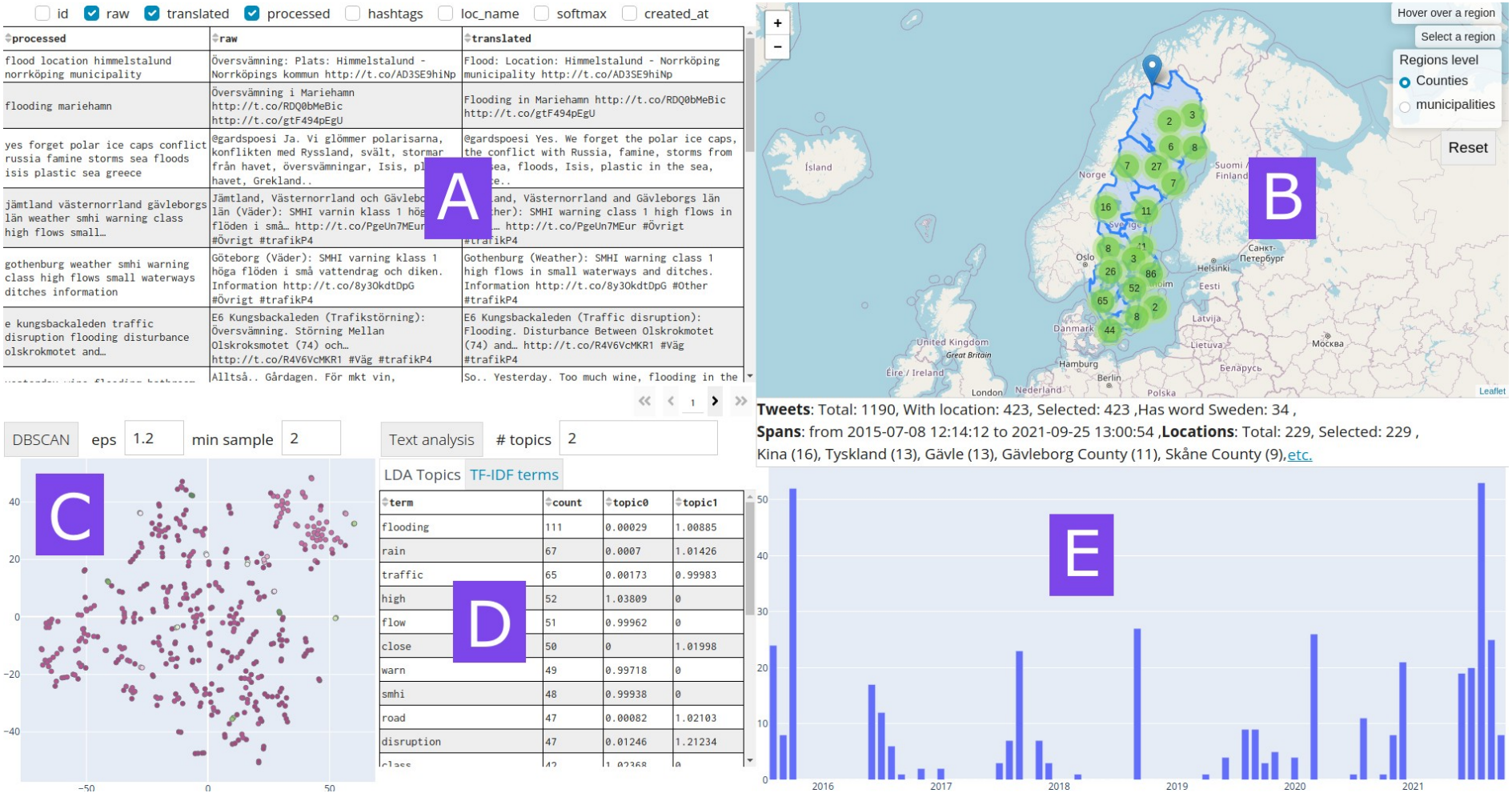


Figure: Visual interface

UPPSALA UNIVERSITET

Figure: Map showing clusters of tweets

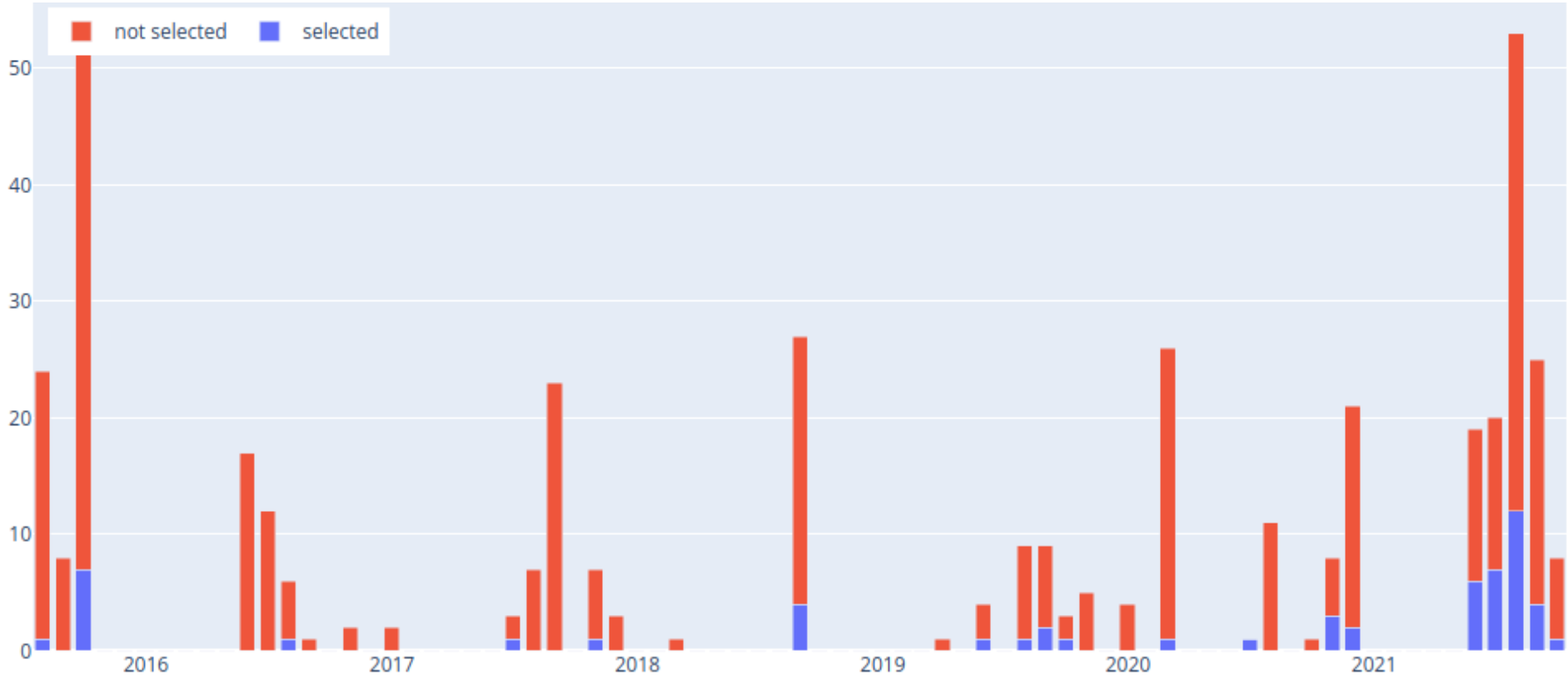


Figure: Histogram for tweets' creation dates

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: Metadata about the tweets

☐ id ☑ raw ☑ translated ☑ processed ☐ hashtags ☐ loc_name ☐ softmax ☐ created_at

processed	raw	translated
seen pictures malmö risk flooding canal central station	@sydsvenskan Har sett bilder från Malmö, är det risk för översvämning vid kanalen vid Centralstationen?	@sydsvenskan Have seen pictures from Malmö, is there a risk of flooding by the canal at the Central Station?
risk landslide county road lot rain p skaraborg	Rasrisk på länsväg 195 efter mycket regn - P4 Skaraborg https://t.co/7Sy16OmjOn	Risk of landslide on county road 195 after a lot of rain - P4 Skaraborg https://t.co/7Sy16OmjOn
e västerås traffic disruption flooding tpl skälbymotet -tpl bäckbymotet direction enköping	E18 Västerås (Trafikstörning) Översvämning. Tpl Skälbymotet (129)-tpl Bäckbymotet (130) i riktning mot Enköping https://t.co/AUwDv7KJyt	E18 Västerås (Traffic disruption) Flooding. Tpl Skälbymotet (129)-tpl Bäckbymotet (130) in the direction of Enköping https://t.co/AUwDv7KJyt
e västerås traffic disruption flooding tpl rocklundamotet -tpl vallbymotet direction örebro	E18 Västerås (Trafikstörning) Översvämning. Tpl Rocklundamotet (132)-tpl Vallbymotet (131) i riktning mot Örebro https://t.co/xUAkMbyRCf	E18 Västerås (Traffic disruption) Flooding. Tpl Rocklundamotet (132)-tpl Vallbymotet (131) in the direction of Örebro https://t.co/xUAkMbyRCf
municipality lidköping large amounts rain caused problems flooding result places lidköping affected need help contact rescue service command center phone	Lidköpings kommun (Övrigt) Stora regnmängder har orsakat problem med översvämning som följd på vissa platser i Lidköping. Är du drabbad och behöver hjälp kontakta Räddningstjänsten ledningscentral på telefon 0510-771719. https://t.co/uMAvCF82xw	Municipality of Lidköping (Other) Large amounts of rain have caused problems with flooding as a result in some places in Lidköping. If you are affected and need help, contact the Rescue Service command center on phone 0510-771719. https://t.co/uMAvCF82xw
västmanland county weather smhi warning class high flows small watercourses	Västmanlands län (Väder) SMHI varning klass 1: Höga flöden i små vattendrag. https://t.co/CVIO7I69gg	Västmanland County (Weather) SMHI warning class 1: High flows in small watercourses. https://t.co/CVIO7I69gg
dalarnas län weather smhi warning class high flows small watercourses	Dalarnas län (Väder) SMHI varning klass 1: Höga flöden i små vattendrag. https://t.co/lobg35tYOV	Dalarnas län (Weather) SMHI warning class 1: High flows in small watercourses. https://t.co/lobg35tYOV
class warning high flows issued small watercourses parts dalarna gävleborg västmanland counties valid wednesday	Klass 1 varning för höga flöden finns utfärdad för små vattendrag i delar av Dalarnas, Gävleborgs och Västmanlands län, gällande onsdag till fredag	Class 1 warning for high flows has been issued for small watercourses in parts of Dalarna, Gävleborg and Västmanland

« ‹ 1 › »

Figure: Tweets' table

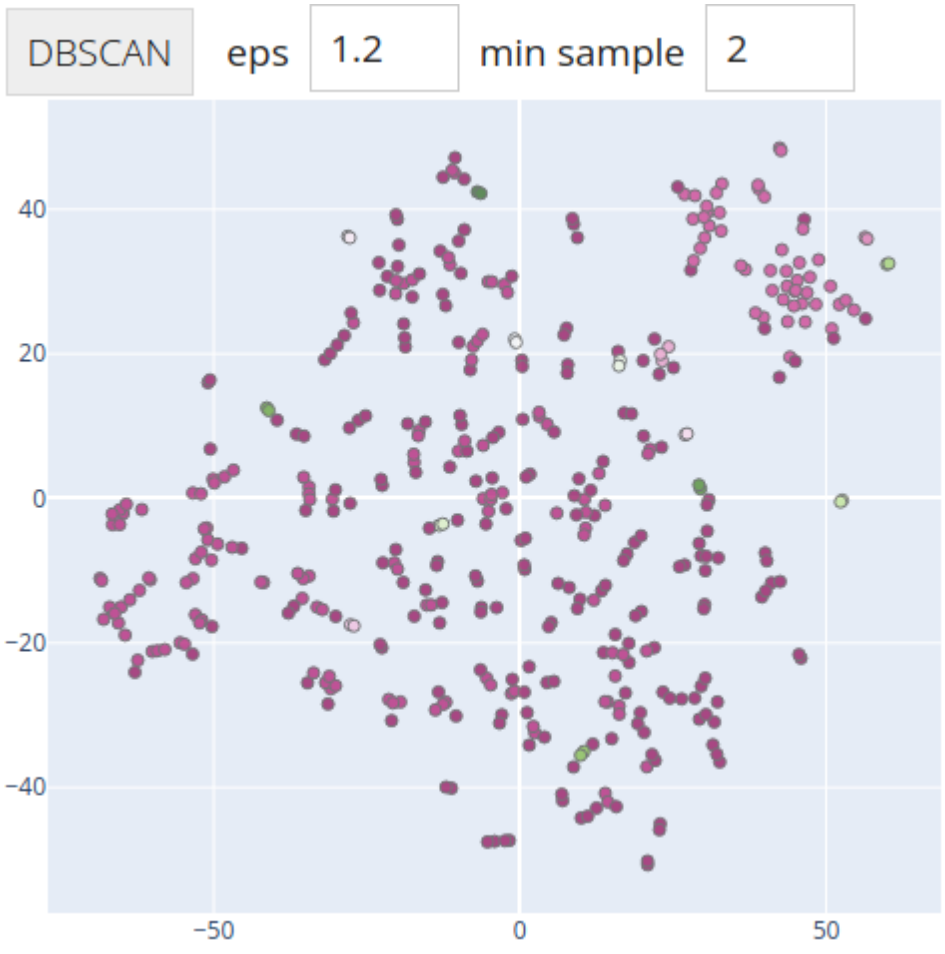


Figure: Scatter plot for t-SNE's space

Text analysis # topics 2

LDA Topics | TF-IDF terms

term	count	topic0	topic1
rain	42	0.02755	0.5921
climate	41	1.24299	0
year	30	1.08646	0
water	27	0.00176	1.03393
traffic	25	0	0.99965
road	22	0	1.18122
disruption	20	0	0.99956
major	18	0.00191	0.83213
change	18	1.1101	0
lv	16	0	1.06206
time	16	1.12201	0

Figure: LDA topic weights

Text analysis # topics 2

LDA Topics | TF-IDF terms

term	mean	count
rain	0.13962	42
climate	0.14282	41
year	0.15428	30
water	0.16512	27
traffic	0.17819	25
road	0.21438	22
disruption	0.19227	20
major	0.27117	18
change	0.2018	18
lv	0.23168	16
time	0.18618	16

Figure: TF-IDF weights

Classifier Performance

- Trained on 20,000 tweets
- The metrics show that the trained classifier is performing well

Table: Evaluation metrics

Accuracy	Precision	Recall	F_1 Score	Confusion Matrix
0.9231	0.8944	0.9181	0.9061	$\begin{bmatrix} 381 & 34 \\ 45 & 568 \end{bmatrix}$

Experiment to validate the pipeline

Extract one week's worth of tweets about a past flood event in Sweden:

- Flood event in Gävleborg and Dalarna counties on 18 August 2021
- 1589 tweets from Twitter API
- 910 left after pre-processing
- 700 classified as flood-relevent
- 247 mentions locations
- 96 mentions Gävle

Misclassified Tweets

Table: Misclassified tweets for floods in Gävleborg and Dalarna

Original tweet	Translated tweet
Blött i Gävle sa Bull.. https://t.co/fV1ChW7ZTR	Wet in Gävle said Bull.. https://t.co/fV1ChW7ZTR
Nån som vet om det är lite blött i Gävle?	Anyone know if it's a bit wet in Gävle?
Att tänka på mycket regn bakåt i tiden o tänka på bl.a. ån i Halland som steg o ställde till det !	Thinking about a lot of rain back in time and thinking about e.g. the river in Halland that rose and made it happen!

Misidentified Locations

- **Original tweet**: Dödssiffran stiger i turkiska översvämningar #**Turkiet** #svpol https://t.co/K6kLRmxQdw
  Identified location: Turkiet, a hamlet[4] in Uppsala county.
  Actual location: Turkey, the country.
- **Original tweet**: Information. Det kraftiga regnovädret över **Gävle** har orsakat översvämningar i arenan. Detta innebär att all verksamhet i Monitor ERP **Arena**, vilket inkluderar bland annat aktivitet på isen samt restaurangverksamheten, tills vidare är pausad. Vi återkommer med mer information. https://t.co/gHDfirq9VS
  Identified location: Årena, an isolated dwelling[5] in Kalmar county.
  Actual location: Gävle.

[4] isolated settlement

[5] consist of not more than 2 households

Visualization (Map)

Tweets: Total: 700, With location: 247, Selected: 247 ,Has word Sweden: 31 ,
Spans: from 2021-08-17 08:15:09 to 2021-08-22 20:44:07 ,**Locations**: Total: 104, Selected: 104 ,
Gävle (96), Tyskland (7), Brynäs (5), Sundsvall (5), Gävleborg County (5),etc.

Figure: Map showing tweets about flood event in Gävleborg

UPPSALA UNIVERSITET

Visualization (Histogram)

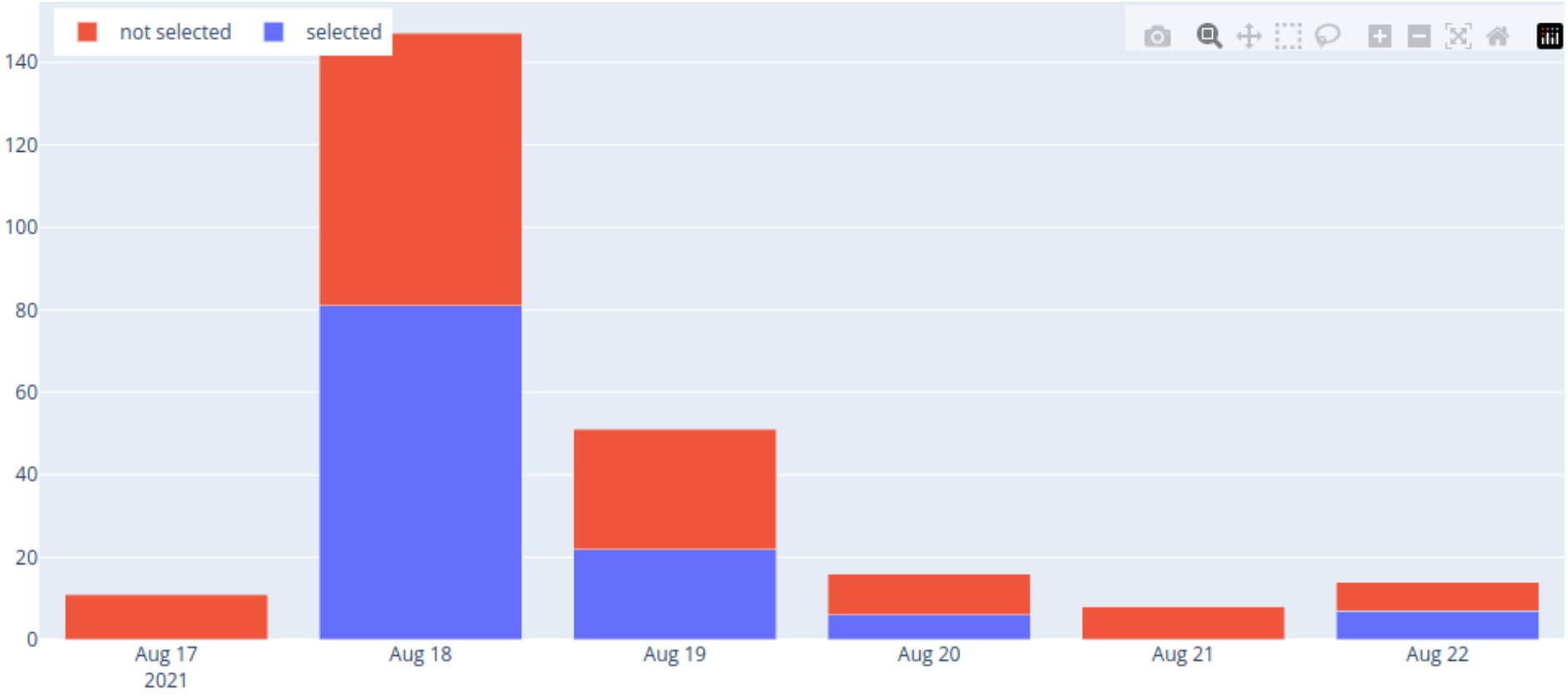
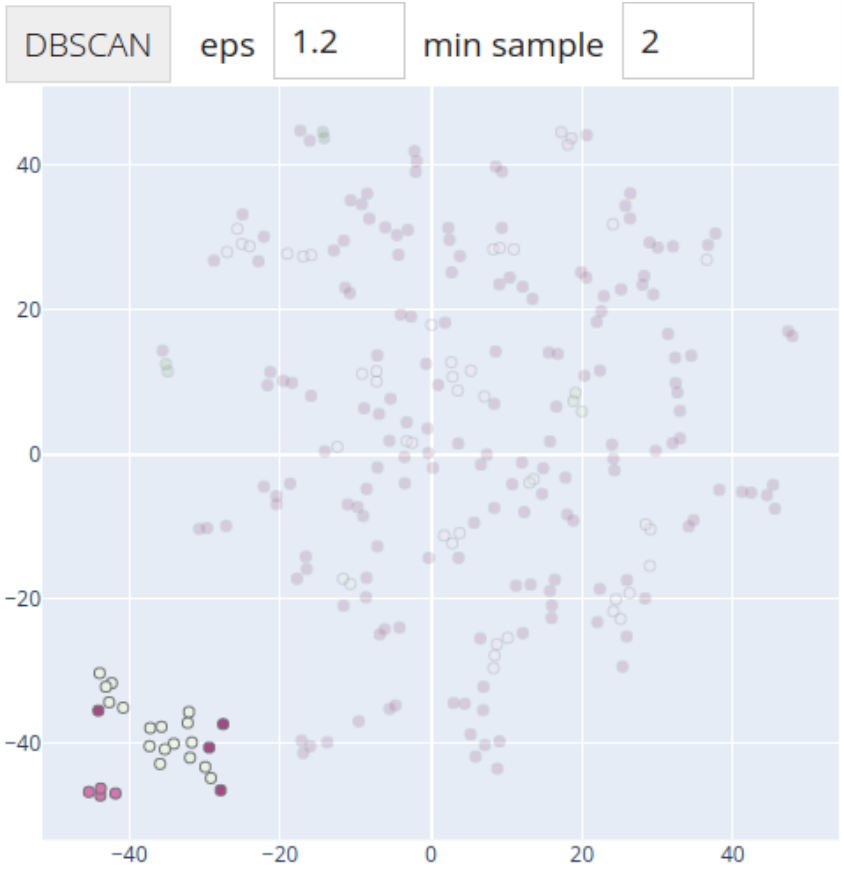


Figure: Histogram showing tweets about flood event in Gävleborg

Visualization (Text Analysis)



Text analysis # topics 2

LDA Topics TF-IDF terms

term	count	topic0	topic1
traffic	23	0.99914	0
disruption	18	0.99898	0
road	15	1.19034	0
lv	15	1.06561	0
flooding	13	0.76406	0
close	5	0.91254	0
water	1	0	0.99007
time	1	0.02118	0

Figure: Scatter plot and LDA table showing a cluster of tweets about flood event in Gävleborg

Short Demo

Results Interpretation, Limitation and Future Word

- Data sources
  - Other social media platforms
  - GDELT
  - Meteorological data
- Classifier performance
  - Process other elements, such as Images and URLs
  - Multilingual support
- Identifying geographical locations
  - Confidence score
  - Handle tweets mentioning several locations
- Visual interface
  - Add more filtering options (e.g. text, LDA topics)
- Future Work
  - Forecasting
  - Other types of disasters

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