

2.1 Data preparation and filtering

Twitter (X)
data

Preprocessing

- a. Keyword filter
- b. Geotag filter
- c. Stopwords, URLs
- d. Named Entity Recognition

*Implementation: NLTK
python package, R regex*

Processed
Tweets

2.2 Content modelling - extracting topics from tweets

Automated Content Model

- a. SBERT Embeddings
(multilingual-MiniLM-L12-v2)
- b. Dimensionality reduction (UMAP)
- c. Clustering (HDBSCAN)
- d. c-TF-IDF

*Implementation:
sentence-transformers, BERTopic
python package*

Topics

Postprocessing Manual filtering

Selected
topics

2.3 Topic interpretation and inter-event comparison

Grouping

Classification framework

Topic
groups

Flood event identification

*Implementation: manual,
changepoint library R*

Event-specific
topic groups

Event-specific
topics

Legend

data
processing

model

output