



**UNIVERSITÉ
DE GENÈVE**



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



POLITECNICO
MILANO 1863



unitar
United Nations Institute
for Training and Research



**Université
de Paris**



Crowd4**SDG**

The Challenge of Collecting and Analyzing Information From Citizens and Social Media in Emergencies: the Crowd4SDG Experience and Tools

Barbara Pernici, Jose Luis Fernandez-Marquez, Carlo Bono, Oguz Mulayin

RCIS 2022, Barcelona May 19, 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944



About us



Barbara Pernici
Professor at Politecnico di Milano
Head of the information
systems group
barbara.pernici@polimi.it



Jose Luis Fernandez-Marquez
Senior Lecturer at University of Geneva
Head of the SDG Accelerator
Crowd4SDG Tech. Coordinator
Joseluis.Fernandez@unige.ch

Carlo Bono (PhD Candidate at Politecnico di Milano)
Oguz Mulayim (Postdoc Researcher at IIIA-CSIC)



Outline

- Round table introduction
- Social media and crowdsourcing for disaster response
 - Challenges + potential role of AI for enhancing the tools
 - Tools and role of AI
- Case study Covid, flood Nepal, flood and gender
- Hands-on Part
 - Visual Cit
 - CSProject builder.
- Discussion





UNIVERSITÉ
DE GENÈVE



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



POLITECNICO
MILANO 1863



unitar
United Nations Institute
for Training and Research



Université
de Paris



Crowd4SDG

Social media and crowdsourcing for disaster response

Jose Luis Fernandez-Marquez

RCIS 2022, Barcelona May 19, 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

2005
announcement of
Pope Benedict XVI

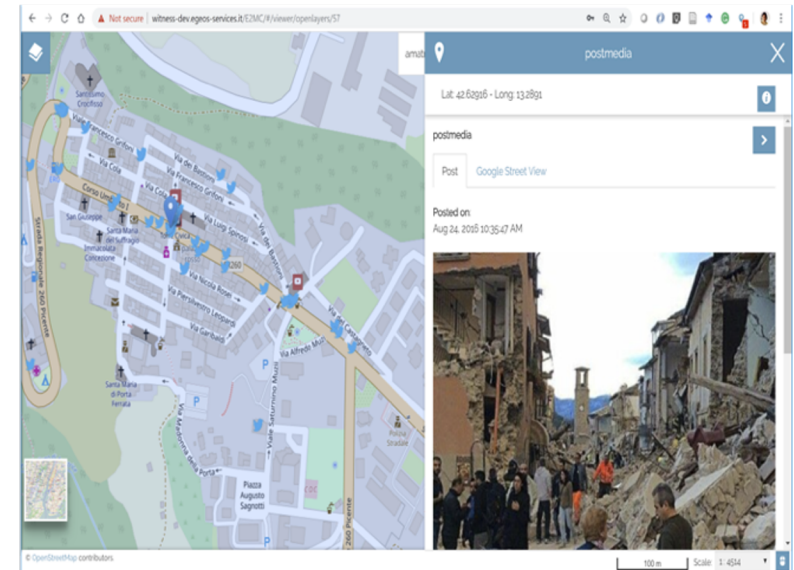
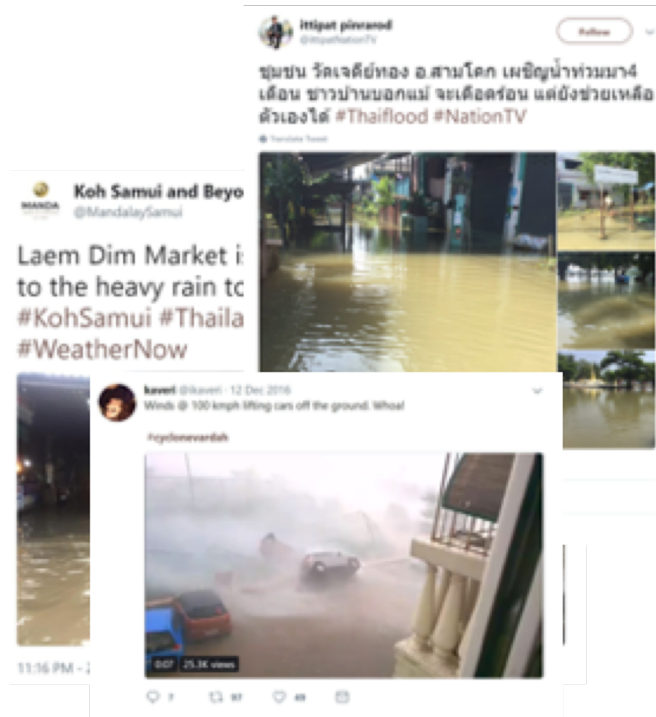


2013 announcement of
Pope Francis





E2mC: Evolution of Emergency Copernicus services





Crowd4SDG

E2mC: Evolution of Emergency Copernicus services

“Every year natural disasters kill around 90 thousands people and affect close to 160 million people worldwide”

Environmental health in emergencies.

WHO



Haiti Earthquake, 2010



Hurricane Katrina, 2005



Chennai floods, 2015

E2mC: Evolution of Emergency Copernicus services

“The first 72 hours after a disaster are crucial; response must begin during that time to save lives” OCHA

Situational Awareness

What is happening and where?



Hurricane Katrina – Satellite Image Corporation

E2mC: Evolution of Emergency Copernicus services



Hurricane Michal – Mexico Beach ESRI



Mexico Beach, Florida, last year after Hurricane Michael made landfall. Douglas R. Clifford / Tampa Bay Times via AP



E2mC: Evolution of Emergency Copernicus services

Main Challenges

- Process of **large volume** of information
- Assessing the **relevance** of the content and the classification
- **Geolocation** of the content gathered from social media
- **Data quality** and reliability.
- **Timeliness**

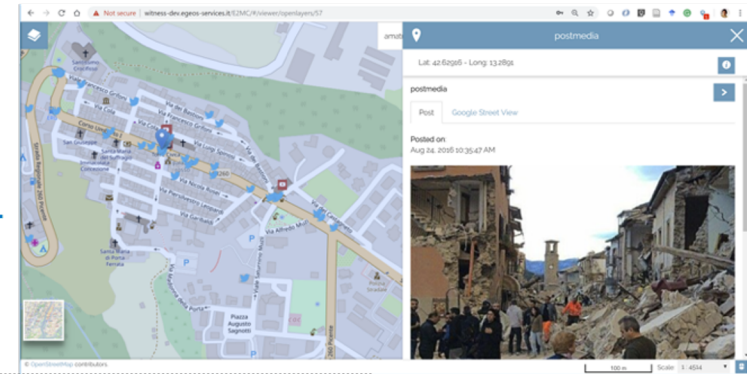


Social & crowd platform

Social Media



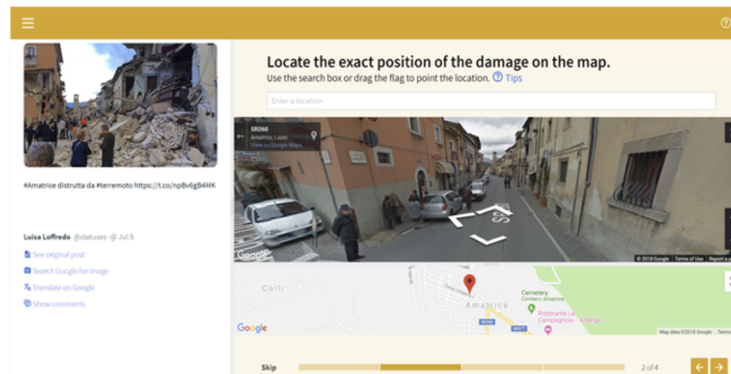
Witness
Component



Relevance



Geolocation



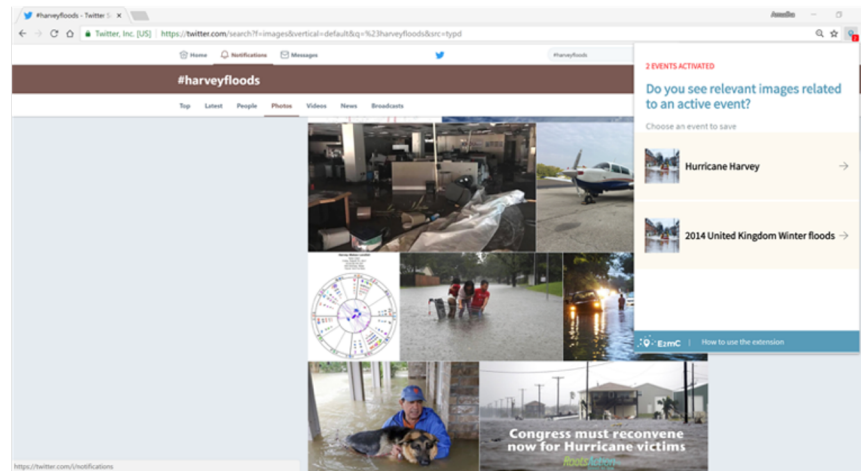
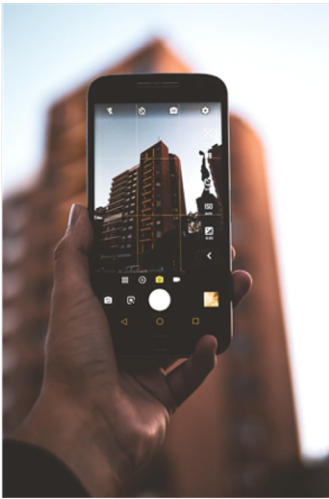
Crowd4EMS

Data Gathering

Crowdsourcing data collection



CROWD4EMS Chrome Extension



Social Media Data

Social Media




Data Analysis - Relevance

← → ↻ 🏠 <https://crowd4ems.org/project/amatricerelevance/task/8710> ☆ ⚙ ⋮

☰ ? Sign in Register

Is this image providing useful information about the Earthquake, Central Italy?



1 post contains this

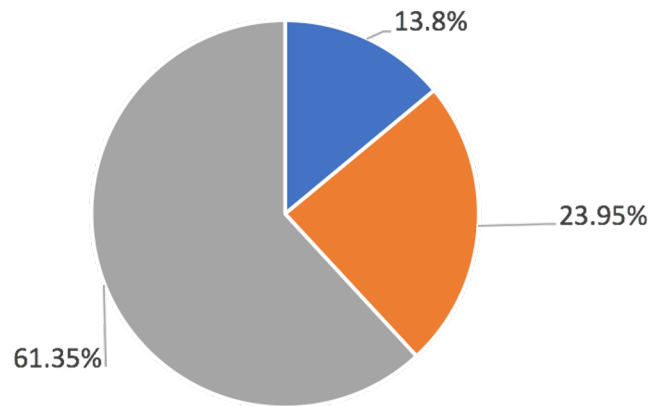
#Amatrice distrutta da #terremoto <https://t.co/npBv6gB4MK>

Luisa Loffredo [More info](#)
tweet / 157 d

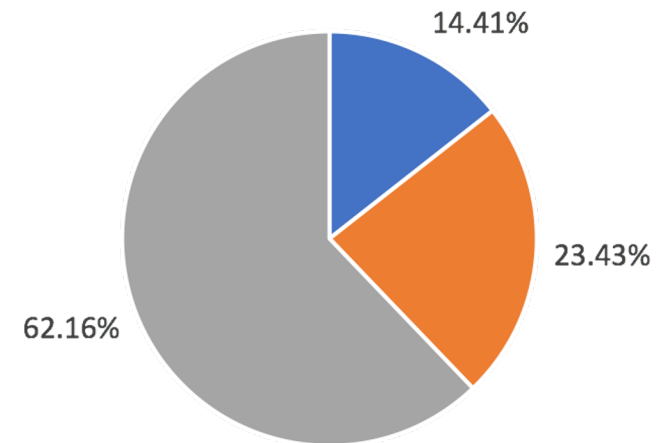
Skip ✗ NOT relevant ✓ Relevant

Data Analysis - Relevance

UK Floods (2014)





Amatrice Earthquake (2016)




■ No Agreement ■ High Agreement ■ Full Agreement



Data Analysis - Geolocation


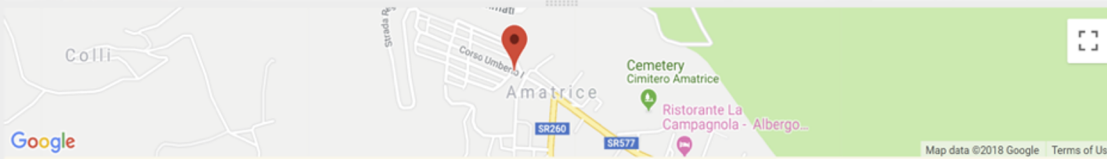


#Amatrice distrutta da #terremoto <https://t.co/npBv6gB4MK>

Luisa Loffredo @statues -@ Jul 5

[See original post](#)
[Search Google for image](#)
[Translate on Google](#)
[Show comments](#)

Locate the exact position of the damage on the map.
 Use the search box or drag the flag to point the location. [Tips](#)

SR260
Amatrice, Lazio
[View on Google Maps](#)

Corso Umberto I
Amatrice
Cemetery Cimitero Amatrice
Ristorante La Campagnola - Albergo...

Map data ©2018 Google Terms of Use

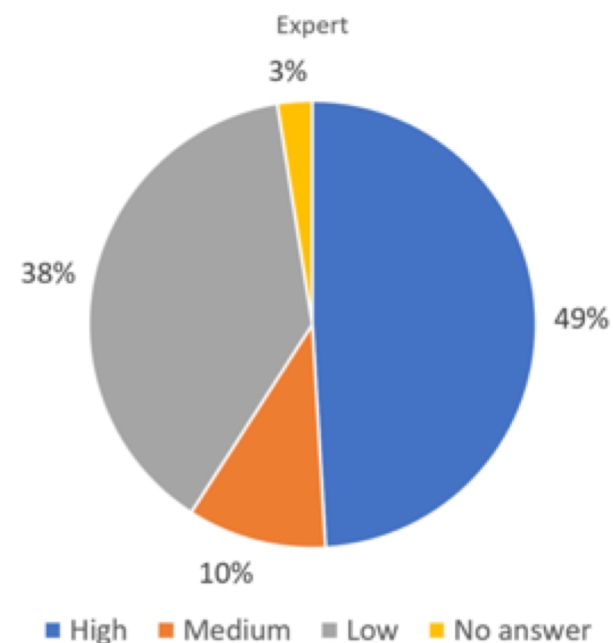
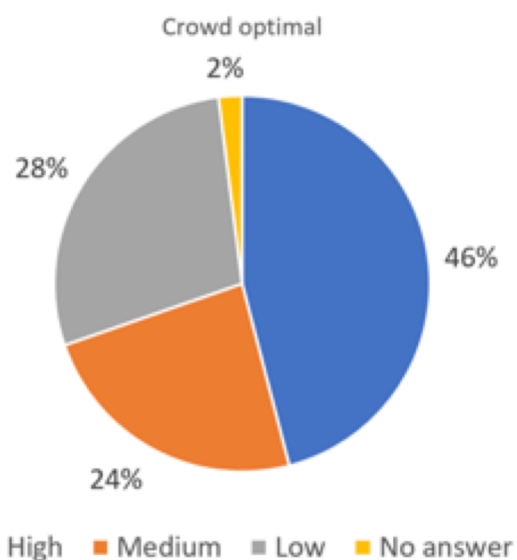
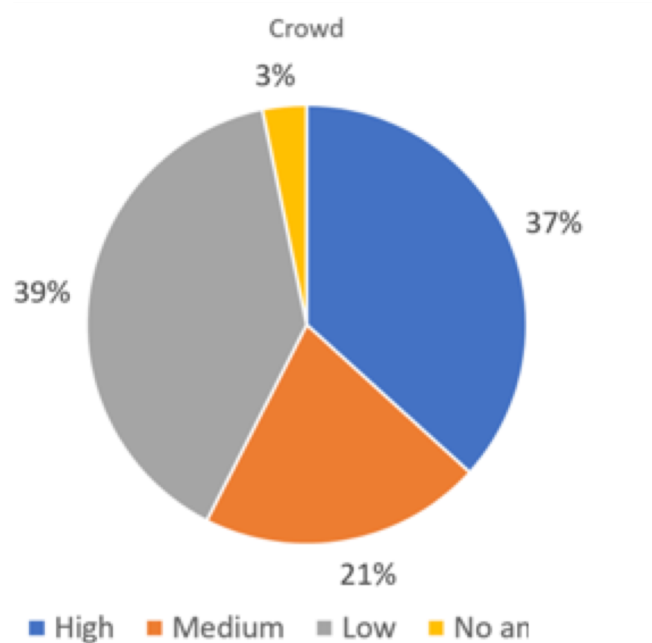
Skip

2 of 4

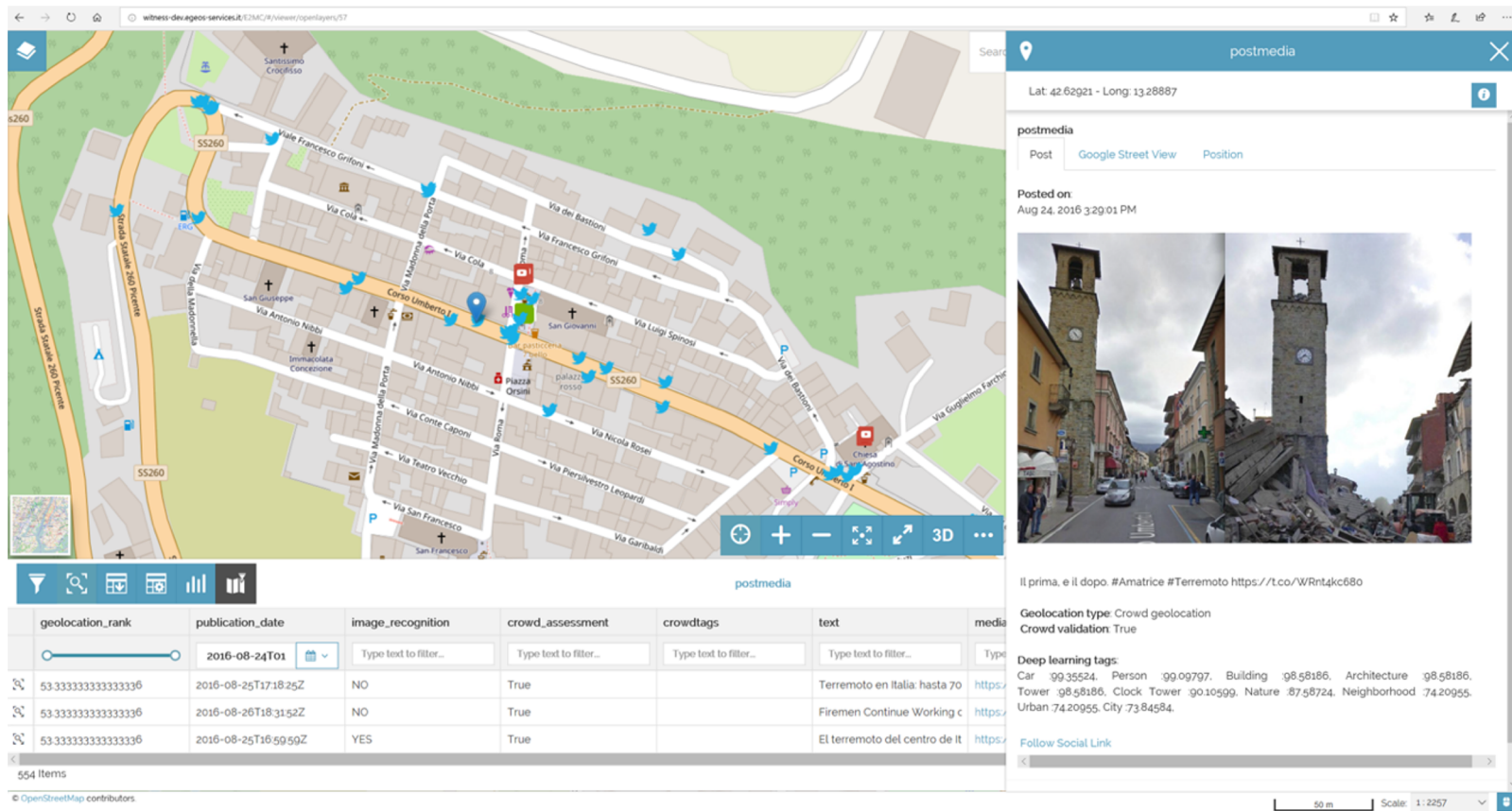
← →



Data Analysis - Geolocation

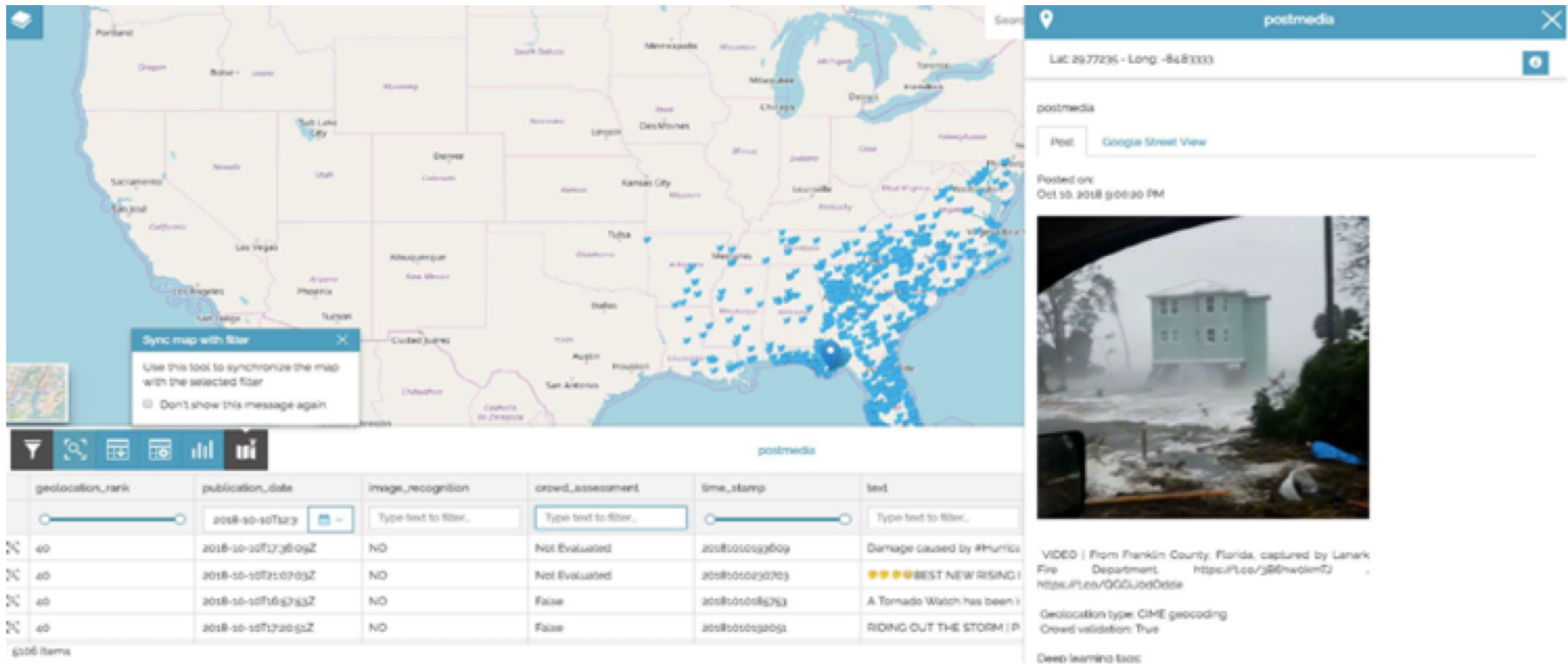


Data Visualization - Witness Component





Data Visualization - Witness Component



Lessons learnt

- Need to improve AI filters to reduce the volume of information before sending that to the crowd
- Geolocation task is still complicated task. It is not a micro task!
- Need to improve the aggregation function to reduce the redundancy needed
- Availability of the crowd
- Engage people beyond data collection and data analysis.





Crowd4**SDG**

<https://Crowd4SDG.eu>



Citizen Science Solution Kit



Tool 1: CS Project Builder - Data Analysis
including CrowdAnalysis component

Tool 2: CS Logger - Data Collection

Tool 3: Visual Cit - Social media data analyzer

Tool 4: Decidim4CS - Open Governance

<https://crowd4sdg.eu/about-2/tools/>



Thank you!

www.crowd4sdg.eu





Crowd4SDG

Consensus models

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Crowdsourced annotations

Majority Voting

	Yes	No
Audi	1.0	0.0
BMW	1.0	0.0
GM	0.0	1.0
Kia	1.0	0.0
Maserati	1.0	0.0
Mazda	1.0	0.0
Seat	0.0	1.0

Probabilistic

	Yes	No
Audi	1.000000	0.000000
BMW	1.000000	0.000000
GM	0.333333	0.666667
Kia	0.666667	0.333333
Maserati	1.000000	0.000000
Mazda	1.000000	0.000000
Seat	0.000000	1.000000

These **do NOT** model individual **annotator** performance.

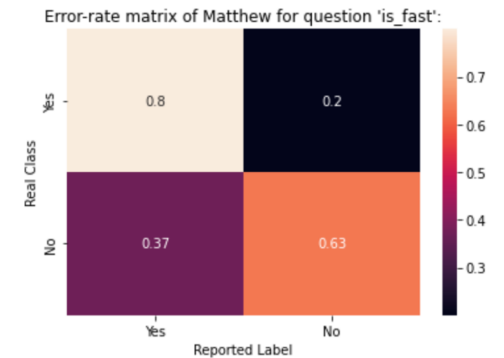
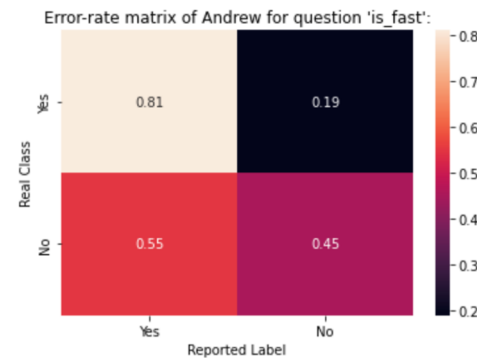
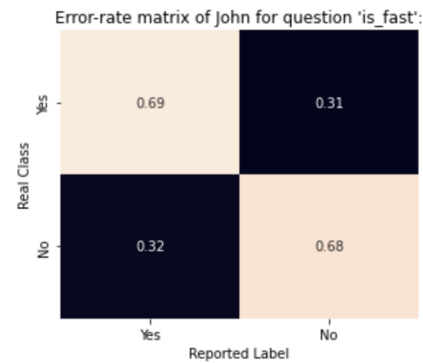


Crowd4SDG

Advanced consensus models - Dawid-Skene

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Annotator error-rates



Consensus

	Yes	No
Audi	0.923616	0.076384
BMW	0.946553	0.053447
GM	0.271625	0.728375
Kia	0.716913	0.283087
Maserati	0.848372	0.151628
Mazda	0.923616	0.076384
Seat	0.096134	0.903866

More reliable consensus is possible thanks to **modelling the annotator behavior**.



Dawid, A. P., & Skene, A. M. (1979). Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics*, 28(1), 20. <https://doi.org/10.2307/2346806>



**UNIVERSITÉ
DE GENÈVE**



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



POLITECNICO
MILANO 1863



unitar
United Nations Institute
for Training and Research



**Université
de Paris**



Crowd4SDG

Case studies

Barbara Pernici

Politecnico di Milano

May 19, 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944



Motivation

- Support emergency response with evidence-based content from social media (focus on Twitter)
- Driver: SDG 13, Climate action
 - Towards improving indicator 13.1.1 Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population





Crowd4SDG case studies

- COVID-19 behavioral impact: **Face masks (VisualCit)**
- **Flood and gender** SDG13 + SDG5
- **TriggerCit: Timeliness** of the delivery in evaluating the **onset** of flood events (Nepal and Thailand)

Building indicators from social media

Social Media provides a trove of information that, if aggregated and analysed appropriately can provide important **statistical indicators to policy makers.**



From images posted to social media

We investigate whether it is possible to obtain such data by aggregating information from **images posted to social media**





Data collection process on Twitter

1. **Focus** on a specific **goal or situation**
1. Twitter sensing with **language-specific** small word dictionaries
1. Selection of **potentially relevant** image contents with VisualCit
1. **Text analysis** of outputs for dictionary refinement
1. **Geolocation** of posts with CIME
1. **Crowdsourcing** to extract relevant information
1. **Visualize results**



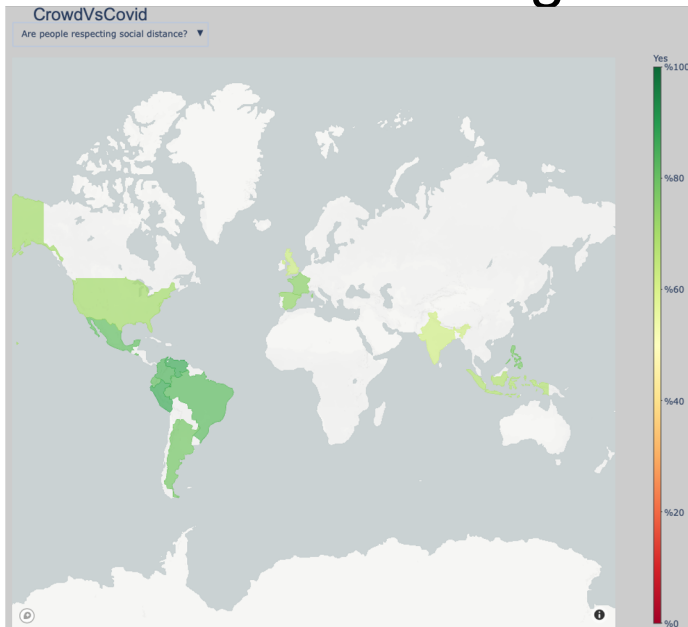
Case studies



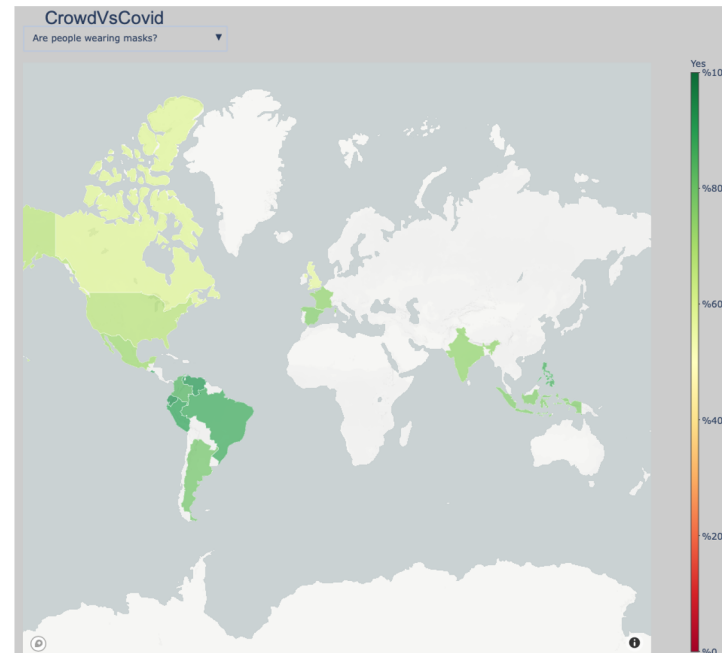


COVID-19 - From Tweets to a Face mask dataset and indicators

- Social distancing



- Face mask usage



Examples end of August, 2020

V. Negri, et al., Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021

Building indicators from social media

Collecting images from social media

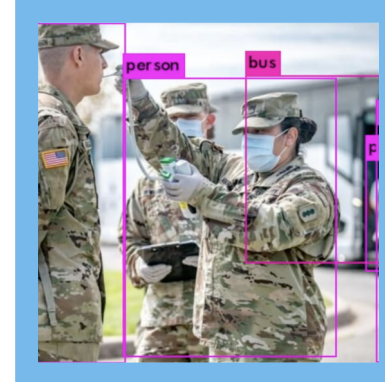


Twitter/social crawling

Keyword selection
(with Citizen Scientists)

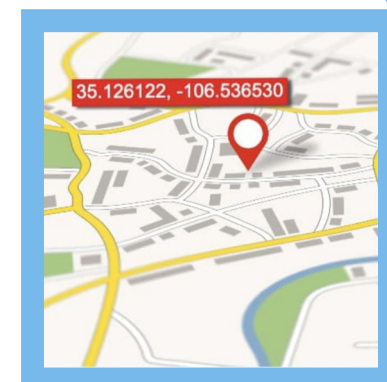


Duplicate URL & Similar Image Filters



Semantic Image Filters

Photo/No photo
Not Safe for Work
Public places
People



Geocoding





This collage features a variety of data visualization techniques:

- Line Graphs:** Three line graphs at the top show trends over time. The leftmost graph has a y-axis from 0 to 100. The middle graph has a y-axis from 0 to 1000. The rightmost graph has a y-axis from 0 to 100.
- Map:** A map of Africa is positioned in the upper center.
- Pie Charts:** Two donut charts are shown. The left one is labeled "50%" and the right one is labeled "24%".
- Bar Chart:** A vertical bar chart is on the left side.
- Area Chart:** A line graph with a shaded area underneath is on the right side.
- Stacked Bar Chart:** A central stacked bar chart uses hexagonal segments. It is labeled "Total population" at the top, "Population by quarter" in the middle, and "Population by year" at the bottom. The segments are labeled with percentages: 20.0%, 30.0%, 20.0%, and 30.0%.
- Line Graph with Markers:** A line graph with circular markers is on the bottom left.
- Line Graph with Area:** A line graph with a shaded area underneath is on the bottom right.



Project Builder

Asking Citizen Scientists

Aggregate Statistics and

- ML Quality evaluation
- Building indicators
- Building new classifiers

Visualisation

<https://pernici.faculty.polimi.it/crowd4sdgpolimi/>

Evaluating the impact of floods on gender equality

- Twitter crawling
- Flood classifier (fine-tuning Xception pretrained with ImageNet)

Is this a flood image?

YES NO

Are there people in it?

YES NO

Are there more women than men?

YES NO CAN'T TELL

Do they seem in danger?

YES NO CAN'T TELL

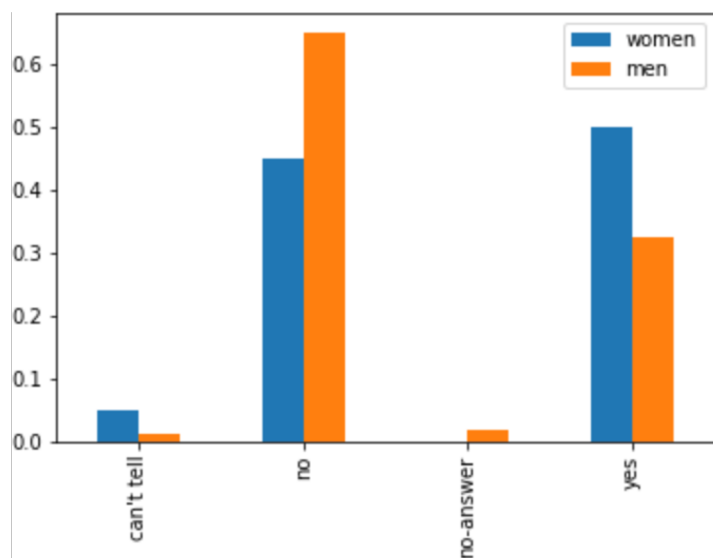
Are there shelters nearby? (Any place giving

The following link to the tweet might help in answering to the questions :

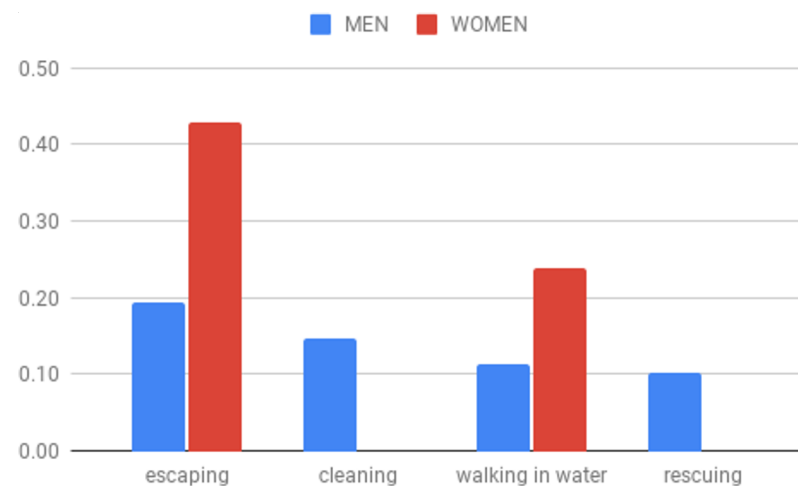
<http://twitter.com/anuuser/status/1391177986897022981>



*Olimpia Rivera, Juan Felipe Calderon, Paul Planchon, Barbara Pernici,
Evaluating the impact of floods on gender equality from social media evidence,
2nd International Research Workshop on Women IS, and Grand Challenges,
Dec. 2021*



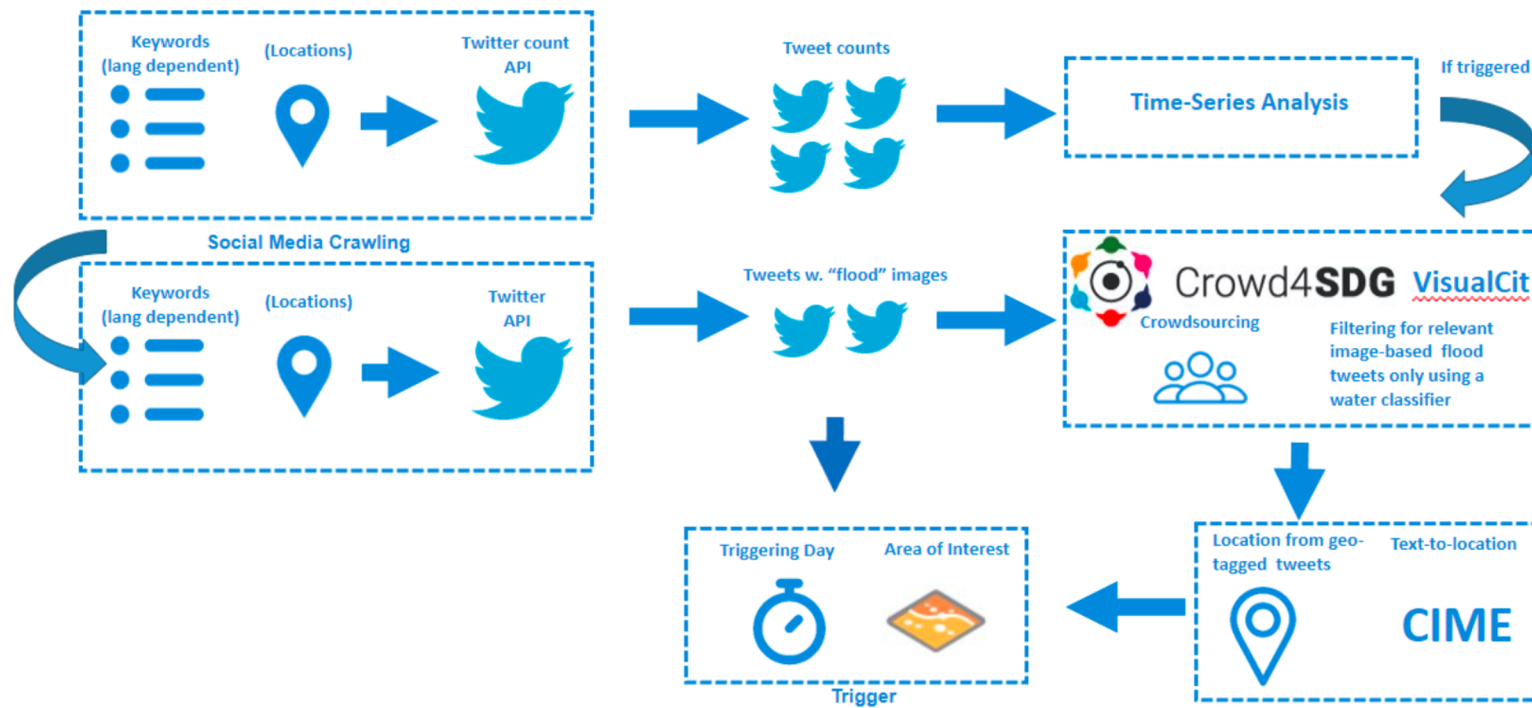
Do they look in danger?



Most popular topics by gender

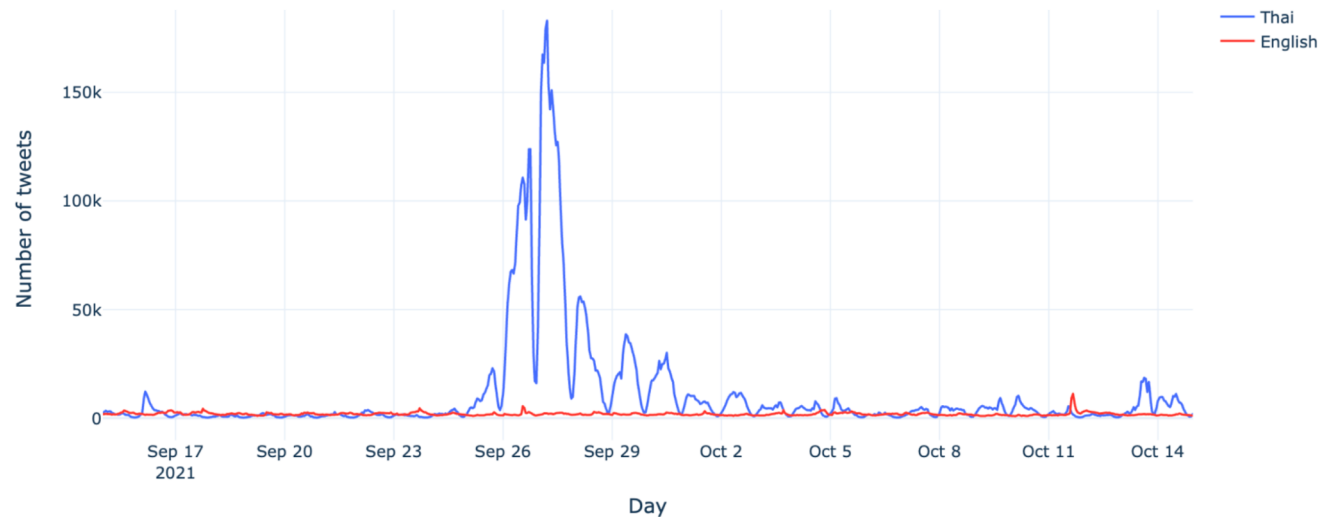
*Olimpia Rivera, Juan Felipe Calderon, Paul Planchon, Barbara Pernici,
Evaluating the impact of floods on gender equality from social media evidence,
2nd International Research Workshop on Women IS, and Grand Challenges,
Dec. 2021*

TriggerCit: Main phases



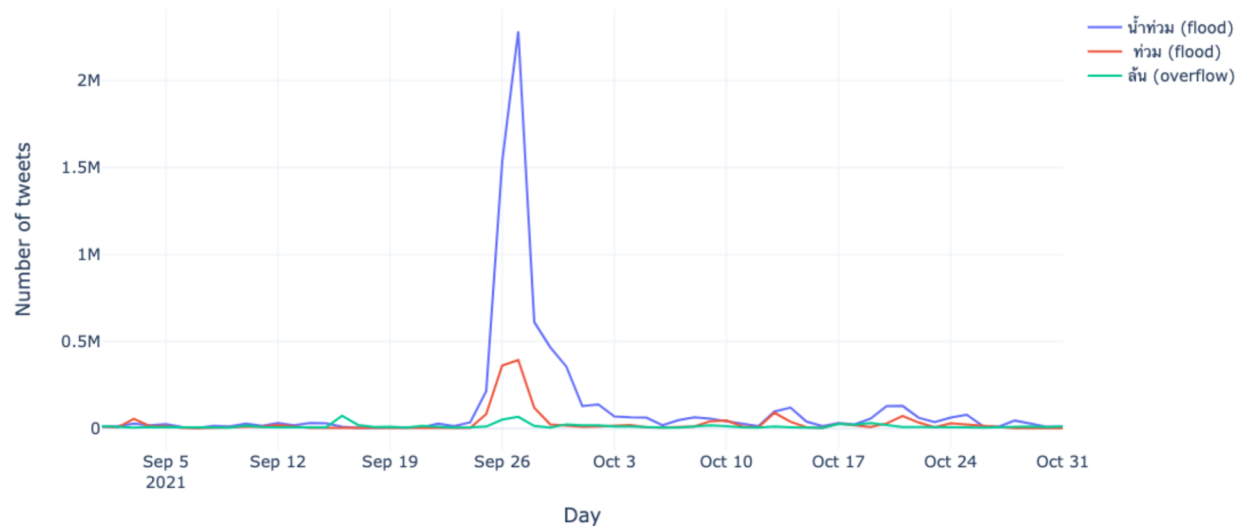
Multilingual context and event detection

We focused on very small, no-cost dictionaries.
Language-based effect of events on social media volumes



Experimental results – Thailand case study

The event onset is clearly identified



Tweet counts for seed dictionary entries

Data	Count
All tweets, September 26-27	4'145'447
No retweets	66'868
Containing images	6'292
Native Twitter locations	227
Overall images	8'774
Passed VisualCit filters	3'056
Places geolocated by CIME	1'671

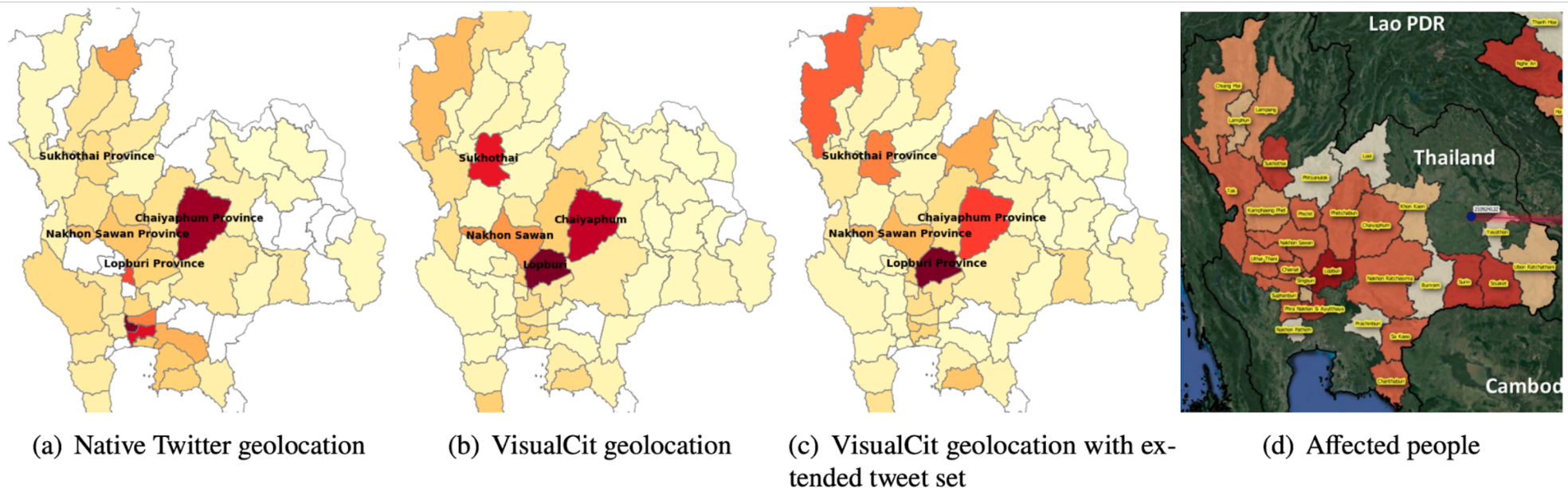
Data volumes through processing steps





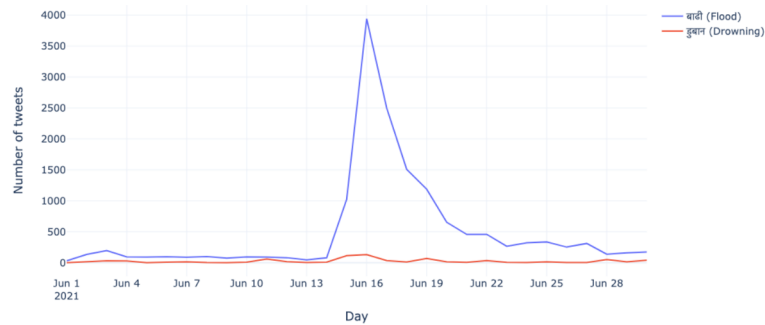
Crowd4SDG

Experimental results – Thailand case study

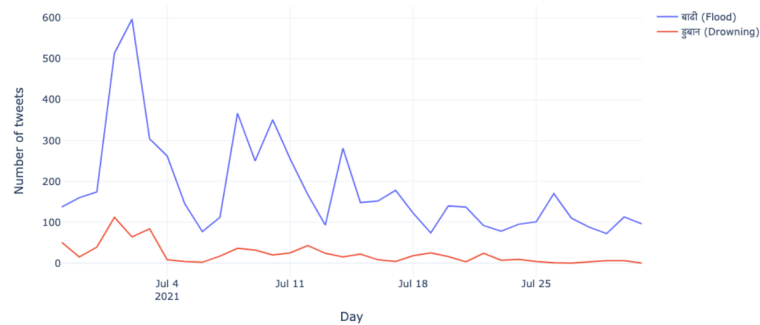


Geolocations / inhabitants ratio by region (a) Twitter native geolocations, (b) Twitter native + CIME geolocated, (c) Twitter native + CIME geolocated from extended dataset (with images + promising text-only tweets), and (d) Number of affected persons by region at September, 28th (source: ReliefWeb)

Experimental results – Nepal case study



(a) Nepali entries in June



(b) Nepali entries in July

Tweet counts for seed dictionary entries

Data	June 16-17	July 1-2
All tweets	6'639	1'225
No retweets	2'807	594
Containing images	261	63
Native Twitter locations	8	10
Overall images	391	80
Passed through VisualCit	218	55
Places geolocated with CIME	51	10

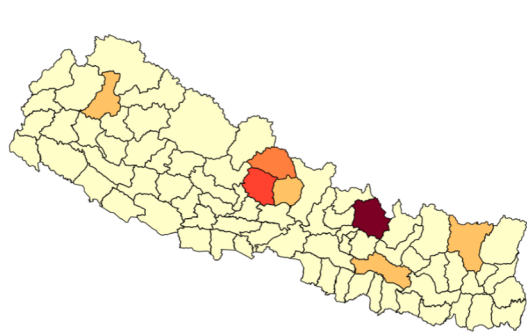
Data volumes through processing steps

The onset of an announcer sub-event is clearly identified

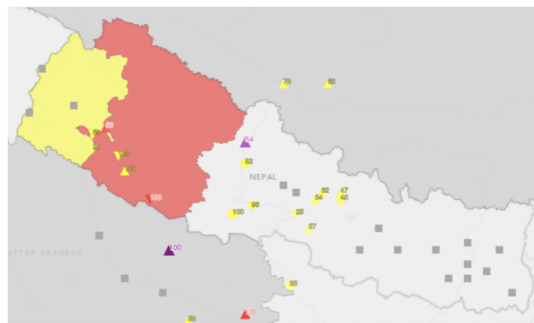


Experimental results – Nepal case study

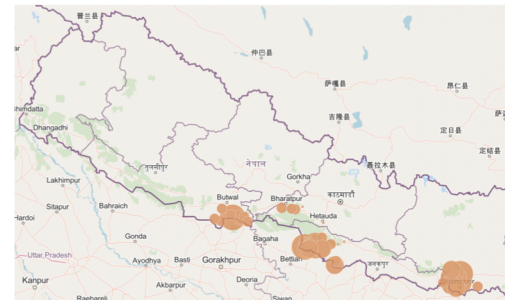
Comparing June 16-17 and July 1-2 events with other sources



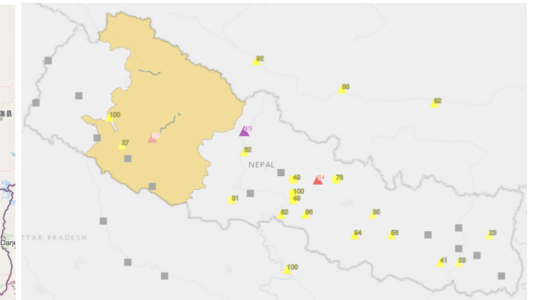
(a) Nepali tweets geolocated with CIME - June 16-17, 2021



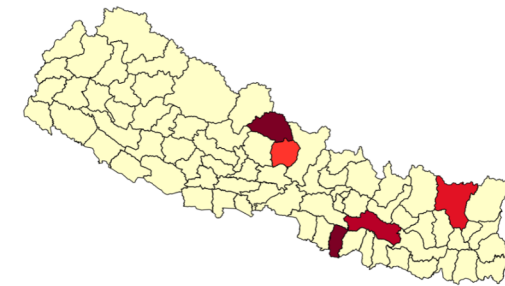
(b) GloFAS reporting points and Rapid Impact Assessment map - June 17, 2021



(a) UNOSAT Flood AI Monitoring Dashboard - July 1-2, 2021



(b) GloFAS reporting points and Rapid Impact Assessment map - July 2, 2021



(c) Geolocated Nepali tweets by district - July 1-2, 2021



(d) Nepal Disaster Risk Reduction portal incident count by district - July 1-2, 2021

Experimental results – Timeliness

Time of the events as reported from different sources

	Dates / Countries	Thailand	Nepal
Activation-based	UNOSAT activation	28/09/2021	30/06/2021
	Copernicus EMS activation	None	None
Manual	GDACS Disaster Alerts	27/09/2021 (Green Alert)	28/06/2021 (Green Alert)
Forecast	GloFAS	24/09/2021	28/06/2021
Manual	FloodList reported news	27/09/2021	04/07/2021
	TriggerCit	26/09/2021	02/07/2021





What we are currently working on

- Use **validated** event data as a ground truth, to benchmark and enhance the architecture components, evaluating uncertainty in space and time
- Explore the robustness of a supervised approach for automated event **triggering**
- Additional for data layers (additional systems, sensors, models, media) for an integrated **multi-modal** approach





Crowd4SDG

References

- Carlo Bono, Barbara Pernici, Jose Luis Fernandez-Marquez, Amudha Ravi Shankar, Mehmet Oğuz Mülâyim, Edoardo Nemni, TriggerCit: Early Flood Alerting using Twitter and Geolocation - a comparison with alternative sources, 2022, accepted for presentation at ISCRAM 2022, May 2022 preprint: <https://doi.org/10.48550/arXiv.2202.12014>
- Carlo Bono, Barbara Pernici. (2022). Twitter dataset of flood-related images for September 2021, Thailand and June/July 2021, Nepal floods [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.6327151>
- V. Negri, D. Scuratti, S. Agresti, D. Rooein, G. Scalia, J. L. Fernandez-Marquez, A. Ravi Shankar, M. Carman and B. Pernici, Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021 [link to dataset and paper](#)
- G. Scalia, C. Francalanci, B. Pernici, CIME: Context-aware geolocation of emergency-related posts, Geoinformatica, 26 (1), 125-157 <https://link.springer.com/article/10.1007/s10707-021-00446-x>



Thank you!

www.crowd4sdg.eu





UNIVERSITÉ
DE GENÈVE



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



POLITECNICO
MILANO 1863



unitar
United Nations Institute
for Training and Research



Université
de Paris



Crowd4SDG

Social media analysis with VisualCit

crawling, extracting information from social media posts

Barbara Pernici, Carlo Bono, Politecnico di Milano

barbara.pernici@polimi.it

RCIS Tutorial, May 19, 2022



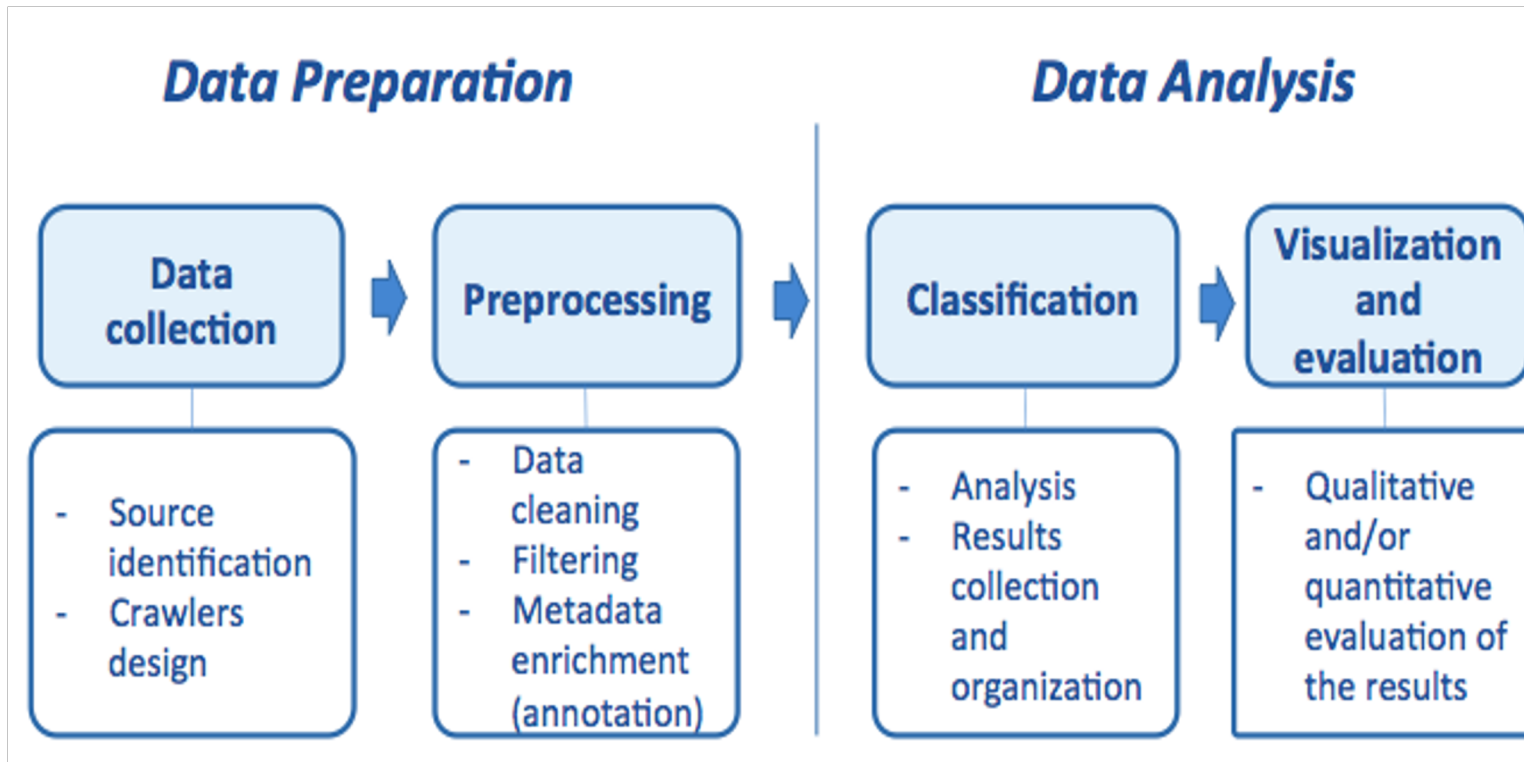
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944



Agenda

- Extracting information from Twitter for emergencies (focus on floods)
- Analysis Pipelines
- VisualCit
- Hands on

Crowd4SDG: VisualCit pipeline



VisualCit Pipeline to build indicators from social media

- Data Preparation
- Collecting images from social media



**Twitter/social
crawling**

Keyword selection
(by Citizen Scientists)

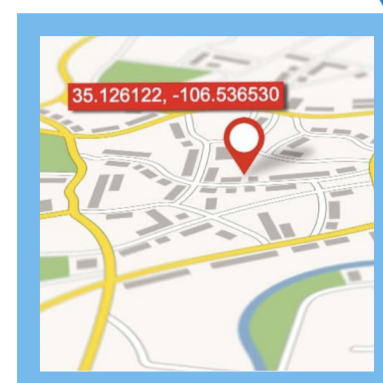


**Duplicate URL &
Similar Image
Filters**



Semantic Image Filters

Photo/No photo
Not Safe for Work
Public places
People

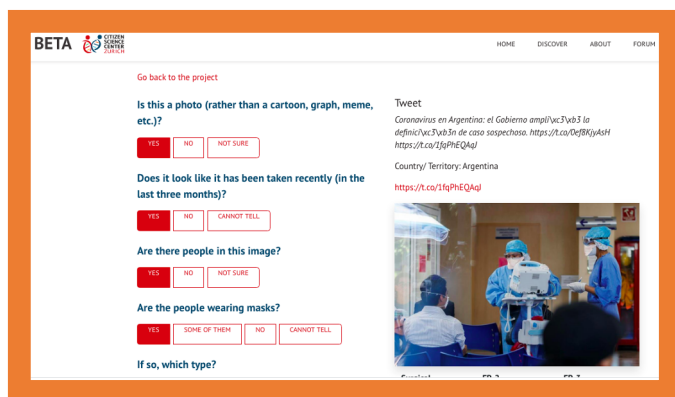


Geocoding

CIME
algorithm



• Data analysis - Deriving indicators



Crowdsourcing

Project Builder

Asking Citizen Scientists

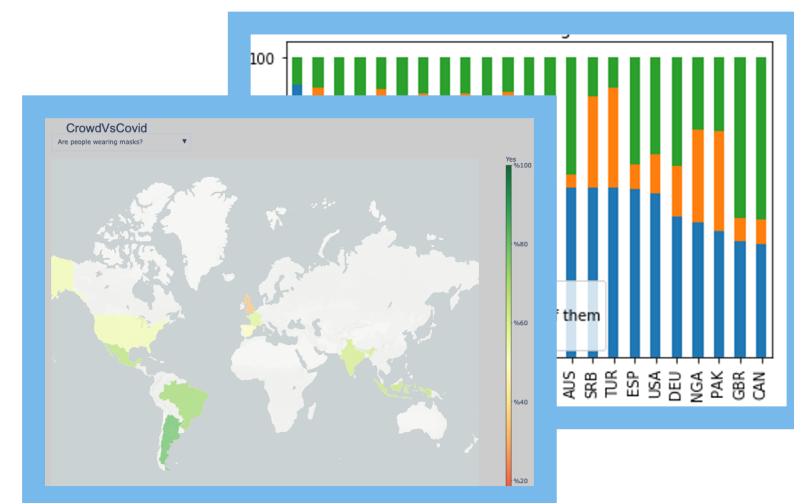


Aggregate Statistics and ML

Quality evaluation

Building indicators

Building new classifiers



Visualisation

Thematic maps

Benchmarks



Building a project with VisualCit

0) Define your goals

1) Select images with VisualCit

- Keywords
- Filters

- Try it:

- <http://visualcit.polimi.it:7778/>

2) Use the collected information

- Set up a crowdsourcing initiative (design questions, identify crowd)
- Compute indicators



← → ↺ Not Secure | http://131.175.120.2:7778



Update

Apps Email - Barbara Pe... Dipartimento di El... 32 egolamento SV SciVal - Publicatio... Cluster 5 - Climat... ep Cisco Webex Mee...

Reading Lis



POLITECNICO MILANO 1863

VisualCit Pipeline Interface

Twitter ▾ Flutwelle OR Hochwasser Number of images: 100 Search



100 items returned Download CSV File

Remove near-duplicate images ▾ Similarity threshold (lower values mean less sensitive): 8 Apply





Crowd4SDG



Crowd4SDG

POLITECNICO MILANO 1863



Pipeline Interface

Twitter ▼ venezia acqua alta Number of images: 100 Search

Venezia

40 items returned Download CSV File

Remove near-duplicate images ▼ Similarity threshold (lower values mean less sensitive): 8 Apply

Venezia

37 items returned Download CSV File

E2mC Keywords for floods in English and Italian

- en=

inundate,inundation,overflow,riverlevel,waterflow,swamped,mudslip,engulf,drainage,flooded,landslide,flooding,riverflow,immerse,torrential,flashflooding,mudflow,floodwater,rainfall,hailstorm,cloudburst,mudslip,precipitation,deluge,deluged,engulfed,flood,landslip,mudslide,submerged,swamp,torrent,waterlevel,tsunami,hail

- it=

alluvione,diga,inondazione,inondazioni,dell'alluvione,torrente,pioggia,piovosità,straripamento,nubifragio,marea,allagamenti,erosione,frana,sommersi,sommersa,bonifica,aumentare,inondato,sommerso,eccezionale,deflusso,precipitazione,diluvio,argine,annegato,salire,terrapieno,allagamento,straripare,palude,drenaggio,valanga,grandine,scolo,lungofiume



VisualCit filters

↑ Remove non-photos ▾ Confidence threshold (%): 90 Apply



43 items returned [Download CSV File](#)

↑ Select event: flooding ▾ Confidence threshold (%): 90 Apply



11 items returned [Download CSV File](#)

↑ Add location information ▾ Apply



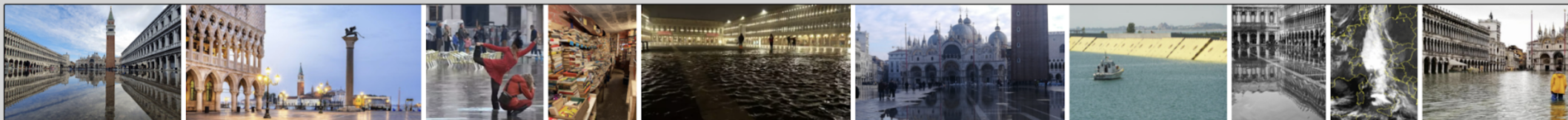


Remove non-photos



Confidence threshold (%): 90

Apply



23 items returned

Download CSV File

Select event: flooding



Confidence threshold (%): 90

Apply



Acqua alta a #Venezia, stamattina 96 cm, 115 cm previsti stasera

#Venice activates flood barriers to keep high tide at bay

Add author locations

#2novembre <https://t.co/lu8S5ZYcDq>
User location: Not initialized

11 items returned

Download CSV File



11 items returned

Download CSV File



Geolocate

Show post locations on a map ▾ Apply



11 items returned

Show fullscreen map

Download CSV File



↑ Add location information ▼ Apply



11 items returned

Show fullscreen map

Download CSV File

-Select- ▼ Apply

Reset Page

Get configuration for this pipeline



Crowd4SDG

VisualCit

Interactive interface: visualization in browser

<http://visualcit.polimi.it:7778/>

Web service: can be invoked inside a Python programme

See [link](#) for details (discussed later)

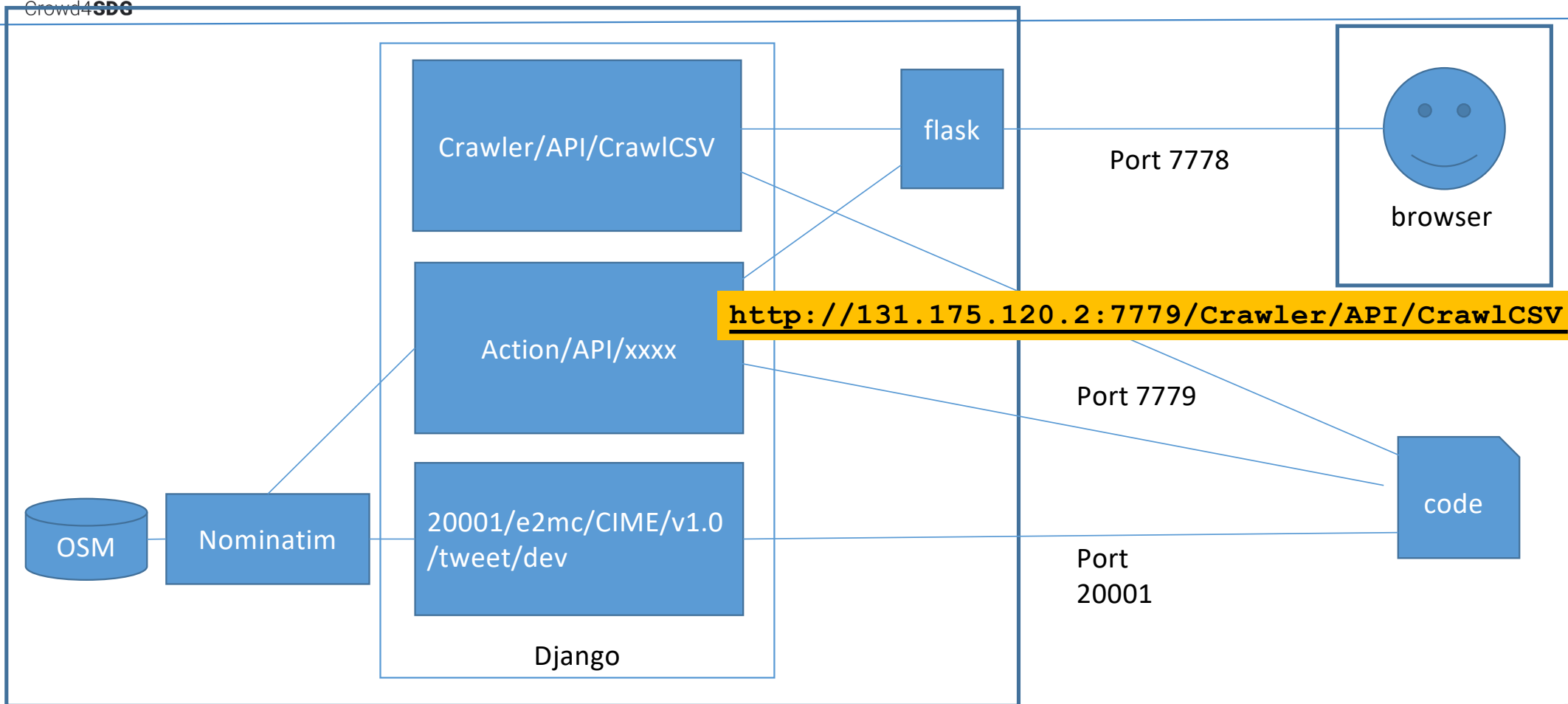
Some **test data** can be found in

[Link to folder](#)





Using VisualCit a service



Server 131.175.120.2



VisualCit endpoints

Endpoint visualization

See services with Command + Option + I on MacOS
Or Command F11 on Windows
then Network

Documentation of available endpoints for VisualCit web services [link](#)

In VisualCit interactive service

Download configuration (Get pipeline configuration)





Crowd4SDG

- Carlo Bono, Barbara Pernici, Jose Luis Fernandez-Marquez, Amudha Ravi Shankar, Mehmet Oğuz Mülâyim, Edoardo Nemni, TriggerCit: Early Flood Alerting using Twitter and Geolocation - a comparison with alternative sources, 2022, accepted for presentation at ISCRAM 2022, May 2022 preprint: <https://doi.org/10.48550/arXiv.2202.12014>
- Carlo Bono, Barbara Pernici. (2022). Twitter dataset of flood-related images for September 2021, Thailand and June/July 2021, Nepal floods [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.6327151>
- V. Negri, D. Scuratti, S. Agresti, D. Rooein, G. Scalia, J. L. Fernandez-Marquez, A. Ravi Shankar, M. Carman and B. Pernici, Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021 [link to dataset and paper](#)
- G. Scalia, C. Francalanci, B. Pernici, CIME: Context-aware geolocation of emergency-related posts, Geoinformatica, 26 (1), 125-157 <https://link.springer.com/article/10.1007/s10707-021-00446-x>





Thank you!

www.crowd4sdg.eu





Details on classifiers

Meme: based on VGG16 (custom)

Flood: based on Xception (custom)

Object: YOLOv5 (also DETR di Facebook, only available in backend)

Scene: PlacesCNN (trained with Places365, based on VGG16)

NSFW: based on MobileNetV2 (not customized)



**UNIVERSITÉ
DE GENÈVE**



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



POLITECNICO
MILANO 1863



unitar
United Nations Institute
for Training and Research



**Université
de Paris**



Crowd4**SDG**

Citizen Science Project Builder

Jose Luis Fernandez-Marquez

RCIS 2022, Barcelona May 19, 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944



CITIZEN SCIENCE TOOLS

CONTRIBUTE TO EXISTING PROJECTS OR CREATE YOUR OWN

The Citizen Science Center Zurich is developing a set of tools that make it easy for scientists and citizens to engage with Citizens Science projects.



Discover



CS Logger



CS Project Builder



Discovering new projects

Discover

Category Thinking ▾

Sort by Newest ▾


32 of 35 Projects in Thinking category



LIFECLEF22 SNAKEID

Humans ID the same snake images as a computer to see how they stack up!

[TAKE A LOOK](#)



VALIDATION OF COASTWARD

Trying to validate my data with CoastWard.org


[TAKE A LOOK](#)




INSTITUTO ANDAPÉ

Projeto de uma calçada ecológica e com menos cimento, assim como o uso de crowdsourcing para pesquisar.

[TAKE A LOOK](#)



Private



Private



Private



Crowd4SDG

Contribute to an existing project

[HOME](#)[TOOLS](#)[DISCOVER](#)[JOSELU ▾](#)[Go back to the project](#)

What Snake is This?

This photo was taken in: **Lusaka - Zambia**

This photo was taken by: **iNaturalist**

Type at least 2 letters for Family, Genus, Binomial or Common Name

[More info](#)SUBMITSKIP

You have completed: **0** tasks out of **150**



Create a new project

Preview Project

Select the type of files that you will work with.



Images



Sounds



Videos



PDF



Tweets



CSLogger



Select the type of files that you will use for your project.

You can choose from multiple options to create your project:

- Images (*.jpg, *.png or *.svg)
- Videos (*.mp4 or *.avi)
- Audios (*.mp3 or *.wav)
- Documents (*.pdf)
- Tweets (@account or #hashtag)
- CSLogger (Integration with mobile app)

For developers: If this template doesn't exactly fit your project, you can try our [expert path](#)



Create a new project

[Go back](#)

What do you want your contributors to do?



Survey



Geo/Survey



Here you can create the survey for your task presenter. You can select between Survey or Geo/Survey

This is a new version!

Two buttons should be visible, if not, please clear the cache and refresh this page.





Crowd4SDG

Create a new project

[Go back](#)

Images Task

Survey

ADD QUESTION



Question 1

☒ Required

Single Choice

Type of question

Question 1

Write your question

The question should not be empty

☐ Option 1

The answer should not be empty

☐ Option 2

The answer should not be empty

ADD ANSWER

CONTINUE

Instructions for Survey:

- Add a question (button "Add Question")
- Choose the type of question (dropdown menu)
- Enter the question and the possible answers
- Set optionality (use "Required" switch)
- Add more answers if necessary (button "Add Answer")
- Add more questions if necessary (button "Add Question")
- If necessary, add "Condition" between questions

Example for Condition:

Question 1: Are the dogs in the image?

- Answer 1: Yes
- Answer 2: No
- Answer 3: I can't say

Question 2: How many dogs in the image?

Condition: Question 2 will be visible only if the answer to Question 1 is "Yes". When on Question 2, click on "Condition", then select Question 1 as conditional question and select "Yes" as conditional answer.

Create a new project - Import tasks

Select your files



Amazon S3 bucket



Dropbox



Flickr



csv File



Select the source of your data.

Your data should be available via url, in particular:

- File hosted on cloud services such as Dropbox, Flickr or Amazon S3 bucket.

Make sure you are authorized to use the data, ie. use free licensed files (Creative Commons or alike), your own files, or copyright files that you are authorized to use.

Our project is created!

[HOME](#)[TOOLS](#)[DISCOVER](#)[JOSELU ▾](#)

RCIS example2

This is an example for the RCIS tutorial

[MODIFY DRAFT](#)[TEST IT!](#)[SHAREABLE LINK](#)[PUBLISH](#)[Info](#)[Statistics](#)[Tasks](#)[Settings](#)

Project data

Project name

RCIS example2

12 characters left

Project short description

This is an example for the RCIS tutorial

80 characters left



Contributing to the project

[Go back to the editor](#)

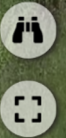
Do you see a flooded area in this picture?

- ☐ Yes
- ☒ No
- ☐ I am not sure

SUBMIT

SKIP

You have completed: 5 tasks out of 24





Managing your project

Info

Statistics

Tasks

Settings

Import Tasks

Import new tasks (files)

IMPORT

Task Presenter

Edit the task presenter (only for experts)

EDITOR

Browse

Check the progress on tasks.

BROWSE

Export Tasks

Export tasks to JSON, CSV, or a CKAN server

EXPORT

Settings

Configure the task scheduler, redundancy, etc.

SETTINGS

Task settings

Task settings

Task scheduler

Change how the tasks are delivered to volunteers

SCHEDULER

Task priority

Change the priority of the tasks

PRIORITY

Delete Tasks

Delete all the tasks and associated answers

DELETE

Tasks Redundancy

Change the redundancy level for the tasks

REDUNDANCY

Export your data

Export in CSV format

Tasks	Task Runs	Results
Export tasks in CSV format	Export task runs in CSV format	Export results in CSV format
EXPORT	EXPORT	EXPORT

Export in JSON format

Tasks	Task Runs	Results
Export tasks in JSON format	Export task runs in JSON format	Export results in JSON format
EXPORT	EXPORT	EXPORT

What's next?



Crowd4SDG

Consensus models

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Crowdsourced annotations

Majority Voting

	Yes	No
Audi	1.0	0.0
BMW	1.0	0.0
GM	0.0	1.0
Kia	1.0	0.0
Maserati	1.0	0.0
Mazda	1.0	0.0
Seat	0.0	1.0

Probabilistic

	Yes	No
Audi	1.000000	0.000000
BMW	1.000000	0.000000
GM	0.333333	0.666667
Kia	0.666667	0.333333
Maserati	1.000000	0.000000
Mazda	1.000000	0.000000
Seat	0.000000	1.000000

These **do NOT** model individual **annotator** performance.

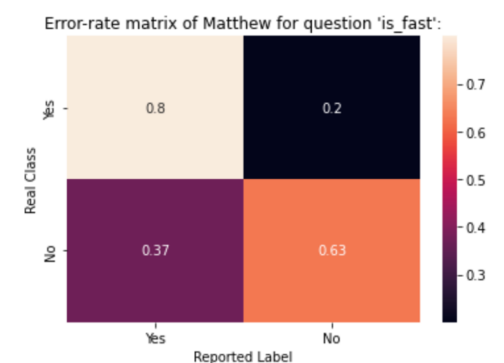
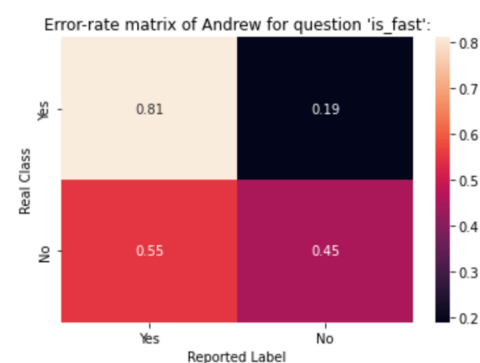
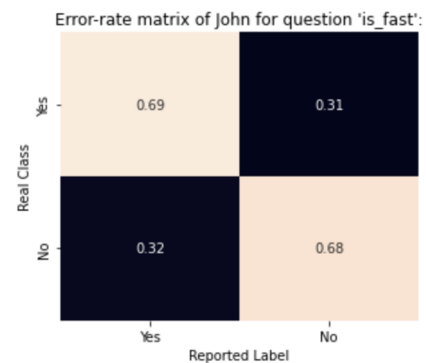


Crowd4SDG

Advanced consensus models - Dawid-Skene

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Annotator error-rates



Consensus

	Yes	No
Audi	0.923616	0.076384
BMW	0.946553	0.053447
GM	0.271625	0.728375
Kia	0.716913	0.283087
Maserati	0.848372	0.151628
Mazda	0.923616	0.076384
Seat	0.096134	0.903866

More reliable consensus is possible thanks to **modelling the annotator behavior**.



Dawid, A. P., & Skene, A. M. (1979). Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics*, 28(1), 20. <https://doi.org/10.2307/2346806>

Search projects

Help Sponsors Log in Register

crowdanalysis 1.1.0

Latest version

Released: Mar 1, 2022

Library to help analyze crowdsourcing results

Navigation

- Project description
- Release history
- Download files

Project links

- Homepage
- Issue Tracker
- Tutorial

Statistics

GitHub statistics:

- Stars: 2
- Forks: 0

Project description

brewing at IIIA-CSIC Tests passing python 3.7 | 3.8 | 3.9 pypi v1.1.0 codecov 95% DOI 10.5281/zenodo.5898579

crowdanalysis

Crowdsourcing Citizen Science projects usually require citizens to classify items (images, pdfs, songs, etc.) into one of a finite set of categories. Once an image is annotated by contributing citizens, we need to aggregate these annotations to obtain a consensus classification. Usually, the consensus for an item is achieved by selecting the most voted category for the item. *crowdanalysis* allows computing consensus using more advanced techniques beyond the standard majority voting. In particular, it provides consensus methods that model quality for each of the citizen scientists involved in the project. This more advanced consensus results in higher quality information for the Crowdsourcing Citizen Science project, an essential requirement as citizens are increasingly willing and able to contribute to science.

Implemented consensus algorithms

- Majority Voting
- Probabilistic
- Multinomial
- Dawid-Skene

crowdanalysis

analyzing crowdsourced data

- Import **annotation data** with preprocessing
- ▢ Calculate **inter-rater reliability**
- ▢ **Model** annotators
- ▢ Compute the **consensus** on annotations
- ▢ Conduct **prospective analysis**
(e.g., 'accuracy vs. number of annotations')
- ▢ **Visualize** consensus & annotator **error-rates**
- ▢ and more ...



<https://pypi.org/project/crowdanalysis/>

👉 [Google Colab Notebook](#) for a hands-on experience



Thank you!

www.crowd4sdg.eu





UNIVERSITÉ
DE GENÈVE



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



POLITECNICO
MILANO 1863



unitar
United Nations Institute
for Training and Research



Université
de Paris



Crowd4SDG

crowdanalysis: *Analyzing crowdsourced data*

Oguz Mulayim

oguz@iia.csic.es

Artificial Intelligence Research Institute (IIIA), CSIC

RCIS Tutorial, May 19, 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944



Crowd4SDG

Consensus models

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Crowdsourced annotations

Majority Voting

	Yes	No
Audi	1.0	0.0
BMW	1.0	0.0
GM	0.0	1.0
Kia	1.0	0.0
Maserati	1.0	0.0
Mazda	1.0	0.0
Seat	0.0	1.0

Probabilistic

	Yes	No
Audi	1.000000	0.000000
BMW	1.000000	0.000000
GM	0.333333	0.666667
Kia	0.666667	0.333333
Maserati	1.000000	0.000000
Mazda	1.000000	0.000000
Seat	0.000000	1.000000

These **do NOT** model individual **annotator** performance.

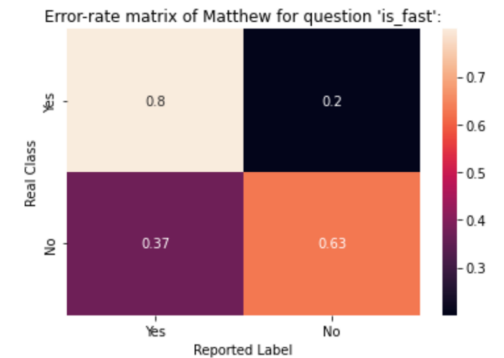
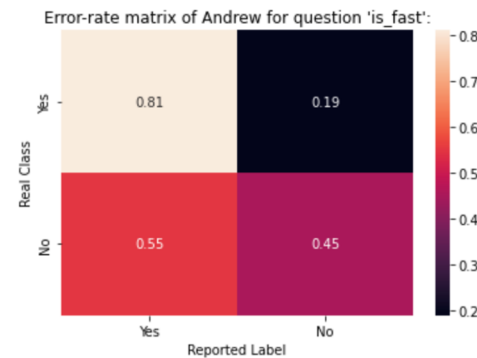
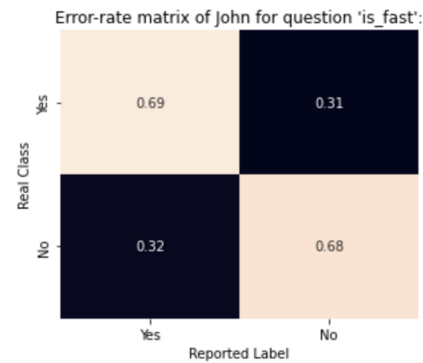


Crowd4SDG

Advanced consensus models - Dawid-Skene

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Annotator error-rates



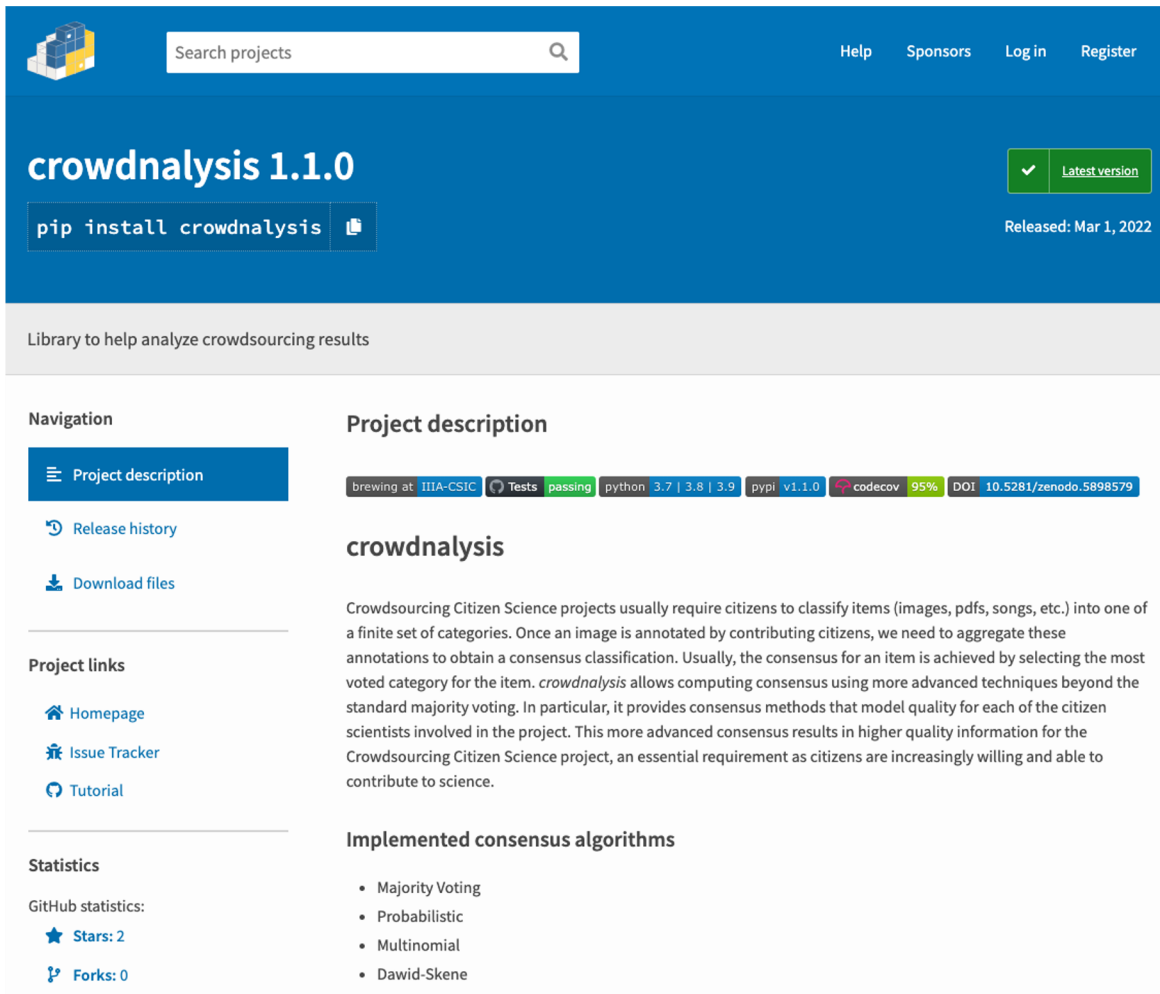
Consensus

	Yes	No
Audi	0.923616	0.076384
BMW	0.946553	0.053447
GM	0.271625	0.728375
Kia	0.716913	0.283087
Maserati	0.848372	0.151628
Mazda	0.923616	0.076384
Seat	0.096134	0.903866

More reliable consensus is possible thanks to **modelling the annotator behavior**.



Dawid, A. P., & Skene, A. M. (1979). Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics*, 28(1), 20. <https://doi.org/10.2307/2346806>



The screenshot shows the PyPI project page for **crowdanalysis 1.1.0**. The header includes a search bar, navigation links (Help, Sponsors, Log in, Register), and a green badge indicating the "Latest version". Below the header, the command `pip install crowdanalysis` is displayed, along with the release date "Released: Mar 1, 2022". The main content area is divided into a left sidebar with navigation links (Project description, Release history, Download files, Project links, Statistics) and a main section for the "Project description". The description explains that crowdanalysis is a library for analyzing crowdsourcing results, specifically for Citizen Science projects. It mentions that it allows computing consensus using advanced techniques beyond standard majority voting. Below the description, a section titled "Implemented consensus algorithms" lists: Majority Voting, Probabilistic, Multinomial, and Dawid-Skene. At the bottom of the page, there is a Python Package Index logo and a URL bar showing `https://pypi.org/project/crowdanalysis/`.

crowdanalysis

analyzing crowdsourced data

- Import **annotation data** with preprocessing
- Calculate **inter-rater reliability**
- **Model** annotators
- Compute the **consensus** on annotations
- Conduct **prospective analysis**
(e.g., 'accuracy vs. number of annotations')
- **Visualize** consensus & annotator **error-rates**
- and more ...



[Google Colab Notebook](#) for a hands-on experience



Thank you!

www.crowd4sdg.eu

