

**computational social media**

**project guidelines & schedule**

**daniel gatica-perez**

**25.02.2022**

# **course project**

defining your project idea

making progress with your project

evaluation & schedule

# defining your project

a topic that you can develop between now and end of semester

teams of 3 people

options

- + your own idea: data analysis (qualitative or quantitative), machine learning, visualization, etc.
- + talk to me immediately if you need ideas

data

- + use publicly available datasets
- + collect your own data

## **examples of previous projects**

**Effect of Fake News in Population Political Polarization: A Brazilian Perspective**

**Differences in Responses to the COVID-19 Outbreak Between Political Spectrums**

**COVID-19 in Switzerland: An analysis of general sentiment using Twitter data**

**Understanding the Political Polarization in Twitter Amidst the COVID19 Pandemic**

**Gender-based Differences in the English-speaking Gaming Community on Twitter**

**Sentiment Analysis in Italian Tweets during COVID-19 Outbreak**

# examples of research datasets

ICWSM data repository (mainly Twitter)

<https://www.icwsm.org/2018/datasets/datasets/>

Appen (formerly Figure Eight, formerly Crowdfunder) data

<https://appen.com/resources/datasets/>

Yelp Dataset Challenge

<https://www.kaggle.com/yelp-dataset/yelp-dataset>

Kaggle datasets

<https://www.kaggle.com/datasets>

Research data sharing platforms like Zenodo

<https://zenodo.org>

# Checklist for your project

## 1. Research questions (RQs)

Pose research questions and tasks

## 2. Experimental protocol

Define target variables

(ground-truth labels)

Define independent variables

(features)

Define subject population

Define data collection process

(questionnaires, APIs, apps)

Address ethics: consent, privacy

Address storage, security, sharing

Get ethical approval if needed

## 3. Data collection

Get datasets according to protocol

Manual: typically low-scale data

Automated: typically large-data

## 4. Processing

Clean, filter, and link raw data

## 5. Coding / Annotation

Enrich raw data with labels

Get external ratings

## 6. Analysis

Check ground-truth quality: reliability

Check feature quality

Compute descriptive statistics

Perform correlation analysis

Apply machine learning (regression,

classification, clustering)

Interpret the results

Provide answers to the RQs

## 7. Communication

Write report

# project schedule & evaluation

## 1. team building

email the list of your team members on **Week 2: Fri 04.03.2022**  
each team will have a designated project mentor

## 2. project pitch

next 2 weeks: discussions of each team with your mentor  
5-minute presentation of your project on **Week 5: Fri 25.03.2022**  
**structure:** title, problem, goals, approach, evaluation

## 3. project progress presentation on **Week 10: Fri 29.04.2022**

5-minute presentation per team about project progress

## 4. final project presentation on **Fri 10.06.2022**

talk: 25-minute presentation + 20-minute questions  
reserve day from 09:00-16:00

## 5. final project report by **Fri 17.06.2022**

ACM conference paper format (6 pages + references + appendix)

**questions?**

[daniel.gatica-perez@epfl.ch](mailto:daniel.gatica-perez@epfl.ch)