

SUPSI

Hackathon II

M-D3030E

Sandra Mitrović, IDSIA
sandra.mitrovic@idsia.ch

Stefano Toniolo, IDSIA
stefano.toniolo@idsia.ch



What is a Hackathon?

*“A **hackathon** is a design sprint-like event in which computer programmers and others involved in software development, including graphic designers, interface designers, project managers, domain experts, and others, collaborate intensively on software projects.”*

«Wikipedia»

Agenda

When		What	Where
Monday 05.02.2024	9:00 - 9:20	Course presentation	Room B1.13
	9:20 - 11:45	Teams working	
	-	Lunch Break	
	13:15 - 16:30	Teams working	
Tuesday 06.02.2024	9:00 - 11:45	Teams working	Room B1.13
	-	Lunch Break	
	13:15 - 16:30	Teams working	
Wednesday 07.02.2024	9:00 - 11:45	Teams working	Room B1.13
	-	Lunch Break	
	13:15 - 16:30	Examination	

Attendance

Attendance is mandatory to pass the class
During all **three** days!!!



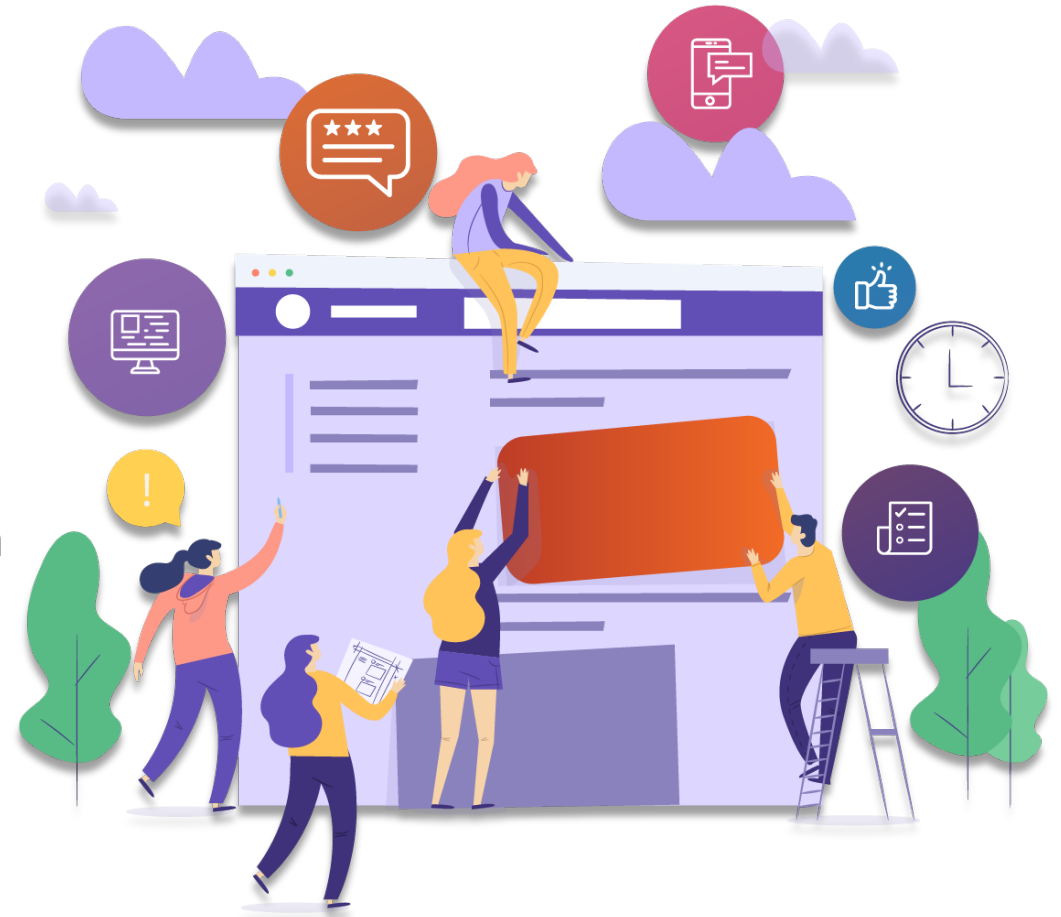
Learning Outcomes

Face a challenge in a short time

Apply your know-how on the field

Organize tasks quickly and efficiently within your team

Thinking out of the box



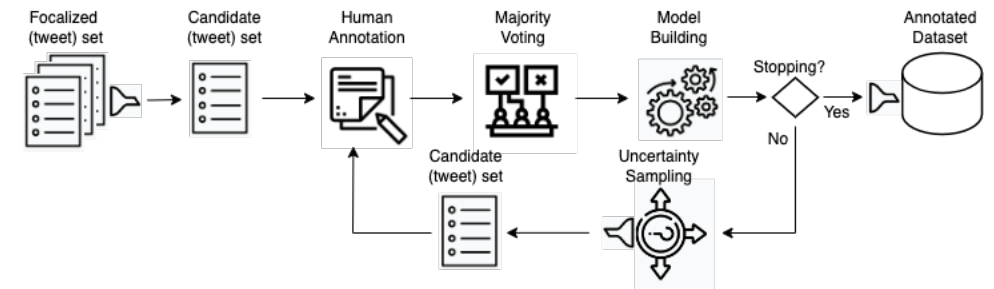
Project theme – Panic

- Mental health is of paramount importance
- Panic disorder (/attacks) understudied in the literature
- Identifying panic from textual form
- Challenges
 - Language is rich
 - Framing panic phenomenon is complex
 - Related to fear but not the same as fear
 - Many physical symptoms shared with other diseases
 - Mentions of “panic”: Panic-related content is not necessarily related to panic attack
 - Lack of annotated datasets



Background

- Tweet annotation project in 2023 (4.8M EN tweets)
- Based on **active learning**
 - Iteratively: Annotate, train & classify, sample most informative
- 6 annotators, 2 annotation rounds, 521 tweet texts considered
 - Annotator disagreement resolved by majority voting
 - “NoMajority” not used for training the classifier
- 3 annotation classes
 - Class “Person Panicking” (denoted as “PP”)
 - a content either describing a personal experience or witnessing a panic attack-like experience
 - Class “Panic - Other” (denoted as “PO”)
 - any content that is somehow related to panic but which does not fall under “PP”
 - Class “Unrelated” (denoted as “UN”)
 - any content not relating to panic, no matter how fearful, anxious or negative it is



Dataset

- Real-life social media data from COVID-19 period
 - From Twitter, original dataset is a Kaggle dataset
 - Starting from 4.8M tweets (3.5M after filtering)
 - Only tweets in English were considered

- Preprocessing: only URLs removed

- Labels

- PP
- PO
- UN
- NoMajority

<p>"Am I the only one that's about to have a nervous breakdown? This sh*t is absolutely crazy and some people are acting as if it's a joke! While my a** up here bout to have a panic attack reading sh*t F*CK IM TIRED OF THIS SH*T!!!! "</p> <p>"I have had 5 panic attacks TODAY thinking I am going to die of #coronavirus can anyone help me what can I do I am so scared I stay in + only go out for food shop but still petrified @DrAmirKhanGP @DrRanj"</p>	Person panicking (PP)
<p>"Consumer panic buying creates supply chain "bullwhip effect" – Read Sarah Rathke's commentary: #supplychain #legal #retail #COVID19 https://t.co/EjZRoxXTfS"</p> <p>"Ihr before PM Lee's public address about taking additional steps to curb Covid19 in SG, the panic buying resumes. Panic buying is a selfish act. Shop sensibly please! #Covid_19 #Singapore"</p>	Pointing at panic buying (PO)
<p>"I think #publicpressure is being manufactured by #media- this information due to its wild inaccuracies will only lead to #panic #covid19 https://t.co/enNpWdACek"</p> <p>"#Covid19. It's bad, for sure. But the media wants you to believe it's the apocalypse, to keep you glued to live news, which is their only means of competing with ad free streaming services. Buy into the caution, but not the panic. 80,000 USA flu deaths in 2017/18 winter btw"</p> <p>"To our media friends all over the world, you have a responsibility to not fuel fear, panic & anxiety. As we report the number of covid19 infections and deaths, please also report the number of people who have recovered. So many people have gone into depression or even worse."</p>	Claiming that media is spreading panic (PO)
<p>"This new situation can be overwhelming, stressful, and intimidating, but don't panic! We have some tips that can keep you crushing it during the COVID-19 crisis. @jcu #StayAtHome #FlattenTheCurve"</p> <p>"WHEN ALL HOPE IS LOST, DON'T GIVE UP! - Short CLIP I am sending love to all the people suffering from mental health problems during the #Coronavirus #lockdown. Don't fear reaching out for help if things go downhill for you . #Itsokaynottobeokay #covid19 https://t.co/wOBccwGtSJ"</p>	Person's plea not to panic and/or to ask for help (PO)
<p>"The captain of a U.S. Navy aircraft carrier facing a growing outbreak of the #coronavirus on his ship was fired on Thursday by Navy leaders who said he created a panic by sending his memo pleading for help to too many people. https://t.co/43O12O8hVb"</p> <p>"I feel half this scare about #coronavirus is created by not-so-experts trying to be one.. half-baked data scientists trying to model the number of deaths and inducing panic in people. I am seeing that some influencers are busy sharing scientific papers on viruses as well."</p>	Pointing at someone else panicking (PO)

- NB: Textual content potentially containing inappropriate speech such as swear words

Dataset

- **Human annotations** → *r1_r2_annotations_liwc_h.xlsx*
 - Consists of:
 - Text (tweet content)
 - Proc_text: text where URLs have been removed
 - A1 – A6: labels provided by 6 annotators
 - Majority_vote: final label decided by majority voting
 - RoundID: ID of the annotation round (1 or 2)
 - Other 119 features generated based on proc_text using LIWC (<https://www.liwc.app/>)
- **Model annotation** → *model_annotations_liwc_h.xlsx*
 - Consists of:
 - Text (tweet content)
 - Proc_text: text where URLs have been removed
 - model_unanimous: label provided by ML model
 - Other 119 features generated based on proc_text using LIWC (<https://www.liwc.app/>)

Tasks and Duties

The next “Tasks and Duties” slides will:

- **describe mandatory tasks to be completed to pass the class (bold text)**
- provide some optional tasks and suggestions to reach the maximum score (plain text)

Pay attention to the timings!



Start from mandatory tasks, with a focus on the optional ones, in order to be able to quickly improve the contents with additional features.

Tasks and Duties – Descriptive analysis

- **Descriptive analysis**
 - Related to human annotation
 - Analysis of *NoMajority* labeled tweets
 - How consistent were the annotators between two rounds?
 - What characterizes each class?
 - Are there any correlations between features and labels?
 - Analysis of annotation for a particular annotator
 - Analysis on the round level
 - Analysis of tweet features and annotation outcome
 - Related to model annotation
 - Are there any correlations between features and labels?
 - ...
 - Human vs. model annotation
 - Compare the two on the benchmark dataset
- **Accompany descriptive analysis with appropriate visualizations**



Tasks and Duties – Unsupervised learning

- **Featurization**
 - Generate TF-IDF vectors for each tweet
 - ...
- **Unsupervised learning**
 - Clustering
- Dimensionality reduction and feature selection
 - PCA
 - Univariate analysis
 - ...
- **Accompany the analysis with appropriate visualizations**



Tasks and Duties – Predictive modelling

- **Predictive model: build a machine learning (ML) model to predict an outcome on the unannotated set**
 - Random forest
 - Gradient boosting
 - Logistic regression
 - Naïve bayes
 - ...

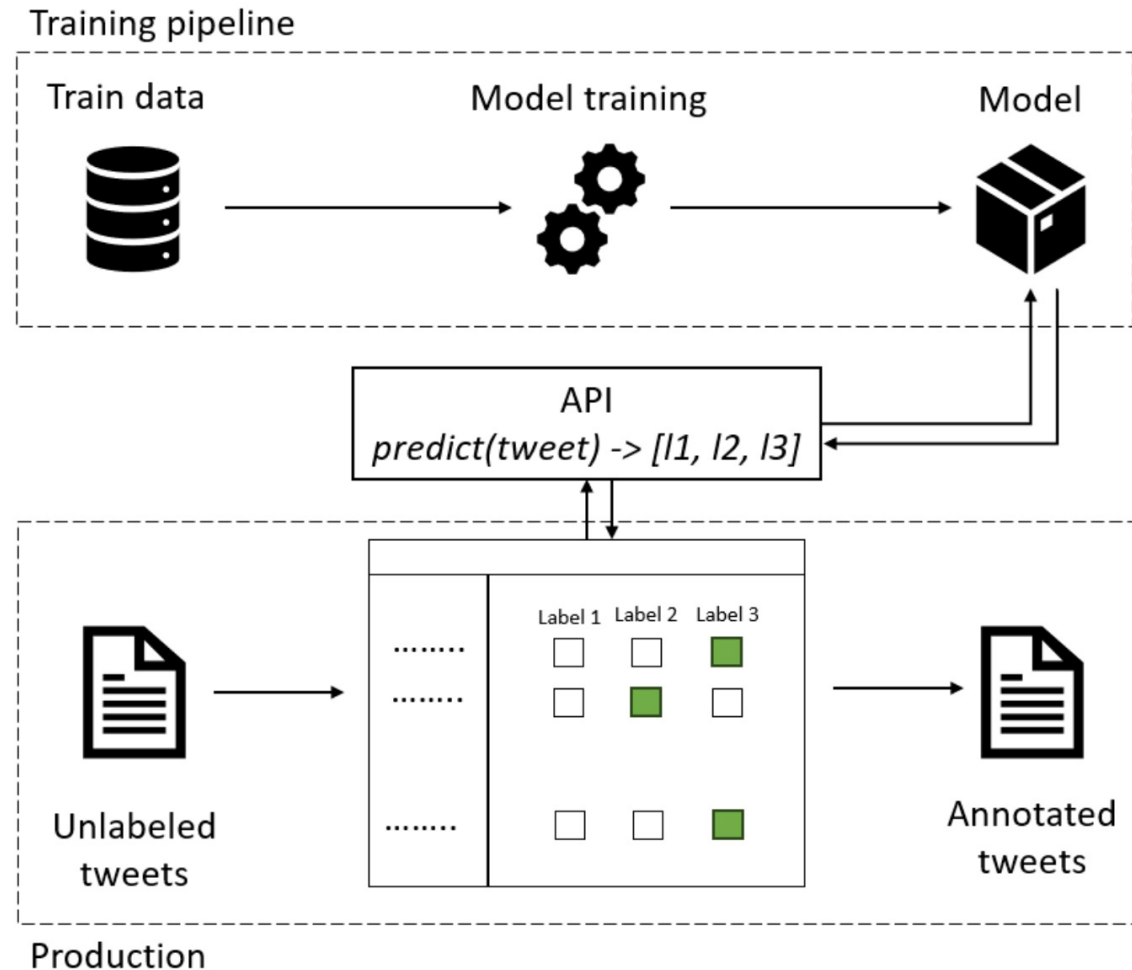
Evaluate model performance; flexibility on train/test dataset

NB: Implement a trivial ML model immediately!!!



Tasks and Duties – Frontend / API

- **Develop an application for tweet annotation**
 - Using e.g. Shiny (or Bokeh)
- **Develop an API to integrate ML model**
 - Wrapper per classifier



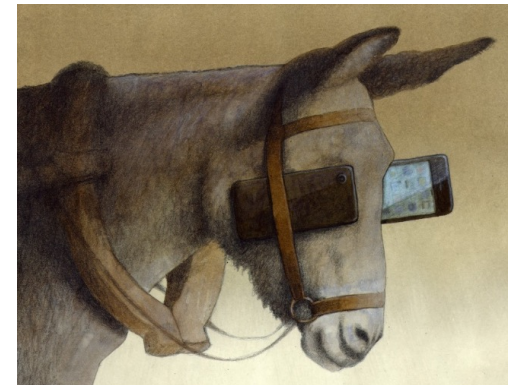
Tasks and Duties – Modules

Teammate cooperation	
Data analysis & data visualization	Frontend / API
Machine learning classifier model	Presentation

Tasks and Duties – Teammate cooperation

Cooperate and interact with your teammates

- Assign roles
- Pay attention to task order
- Choose the main topics together
 - Project timeline
 - Technology
 - Interfaces
 - Libraries
 - ...
- Do not use blinkers, you are a team



Source: https://img.cdn-pictorem.com/uploads/collection/P/PAWEL/900_Pawel-Kuczynski_blinkers.jpg

Tasks and Duties – Presentation

- 15-20 minutes each group
- Each member presents his contribution
- Cite your sources
- Upload the following material in your MS Teams channel before your examination session
 - Presentation (pdf)
 - Source code (zip)
 - Analysis outcomes (pdf)
 - Any further material (zip)
- Live demo



Tips

- Understand data and start by working on a small subset
- First develop a trivial ML model

Useful links

- Scikit learn documentation: <https://scikit-learn.org/stable/>
- LIWC:
 - Tool website: <https://www.liwc.app/>
 - Some papers:
 - LIWC-22 [paper](#)

Enjoy!

