

The Excellent Biological Investigators present...

SentimentHub

Sam Horsfield, Leonie Lorenz, Jacqueline Toussaint



SentimentHub
by the Excellent Biological Investigators

Team roles



Samuel Horsfield

Pipeline development and implementation



Leonie Lorenz

Data analysis and model development

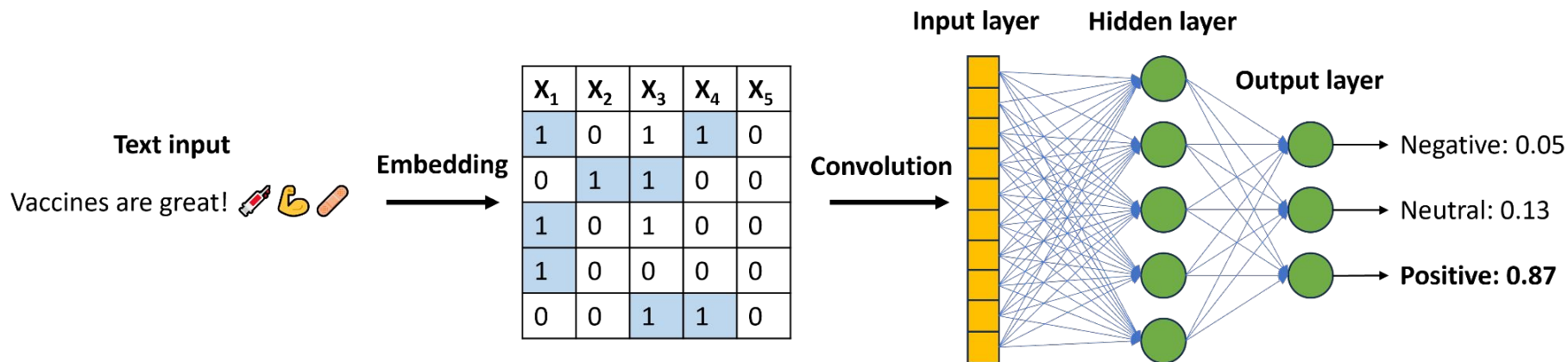


Jacqueline Toussaint

Website development & design

What is Natural Language Processing (NLP)?

1. Takes unstructured text input.
2. Transforms into a structured matrix (**embedding**).
3. Further transforms into 1D array (**convolution**).
4. Generates output using neural network (**hidden and output layers**).



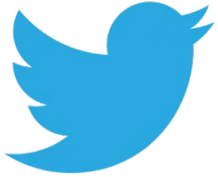
How can NLP be applied in public health?

- Social media sites provide rich data source to understand public sentiment towards public health interventions.
- Sentiment data provides an alternative datasource for predictive modelling, useful when epidemiological data is lacking.
- Twitter datasets have been used previously to:
 - Identify trends in sentiment towards different vaccine types (Marcec & Likec, 2022)
 - Correlate tweet sentiment with vaccine uptake (Qorib *et al.*, 2023)
 - Identify and geolocate increases in public-health keyword use (Espinosa *et al.*, 2022)
- We can use NLPs to predict sentiment, most likely location of user and identify misinformation.

Analysis Pipeline

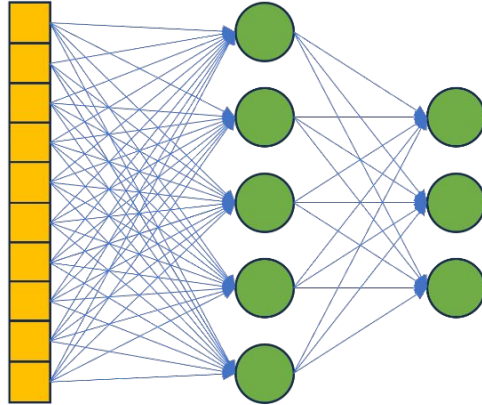
- Available:
 - Sentiment and location NLP, temporal analysis, negative sentiment warning system

Social media API



Credit: Twitter

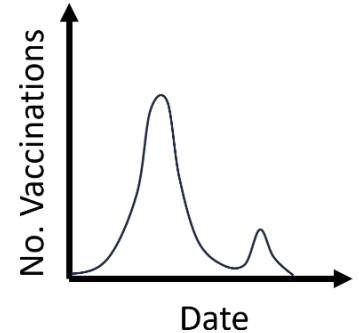
Sentiment and location NLPs



Temporal/Geographical analysis

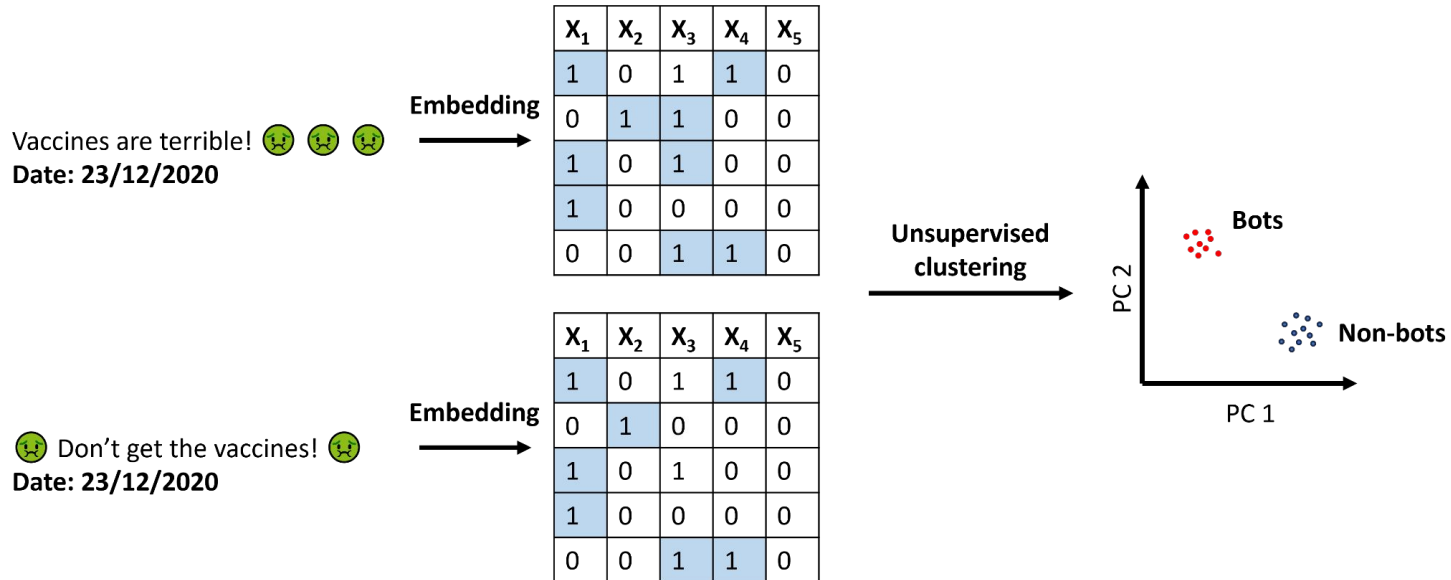


Modular analysis

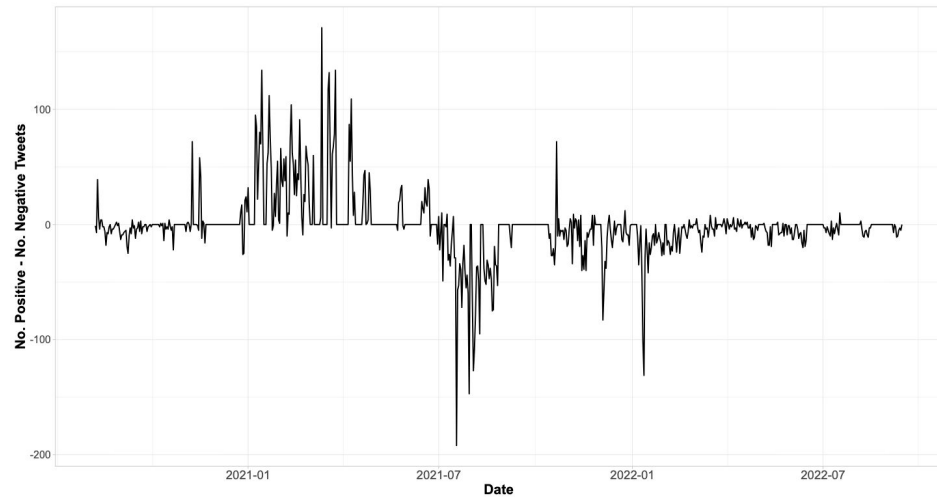
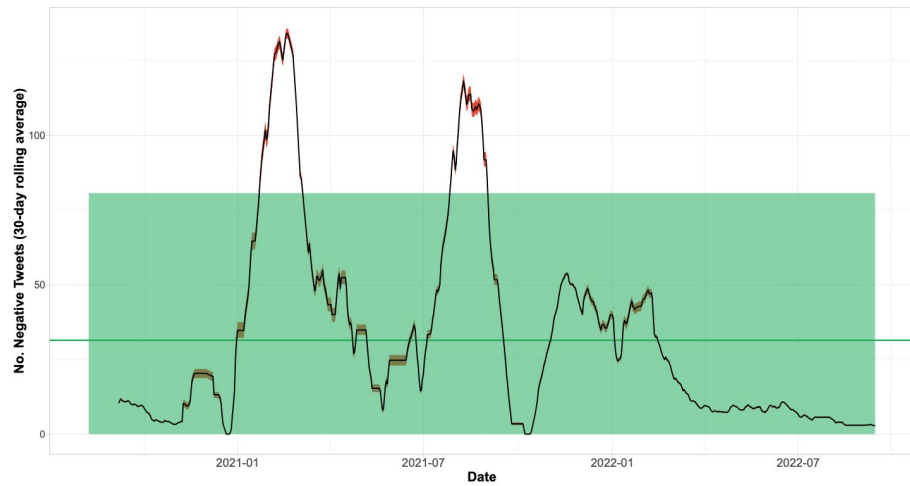


Identification of misinformation (coming soon)

- Identifying misinformation using date, sentiment embedding and unsupervised clustering - doesn't require training.
- Validate manually and using training datasets e.g. Hayawi *et al.*, (2022)



Vaccine Sentiment



Relating sentiment to vaccine uptake

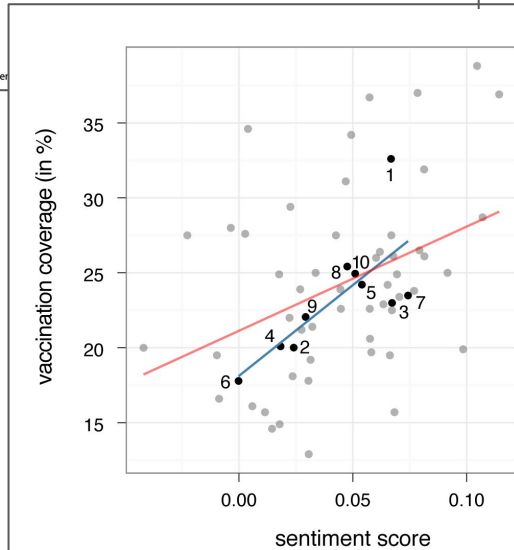
OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control

Marcel Salathé*, Shashank Khandelwal

Center for Infectious Disease Dynamics, Department of Biology, Penn State



Research | Open Access | Published: 02 September 2022

Factors associated with COVID-19 vaccine intentions during the COVID-19 pandemic; a systematic review and meta-analysis of cross-sectional studies

Emily Terry ✉, Sapphire Cartledge, Sarah Damery & Sheila Greenfield

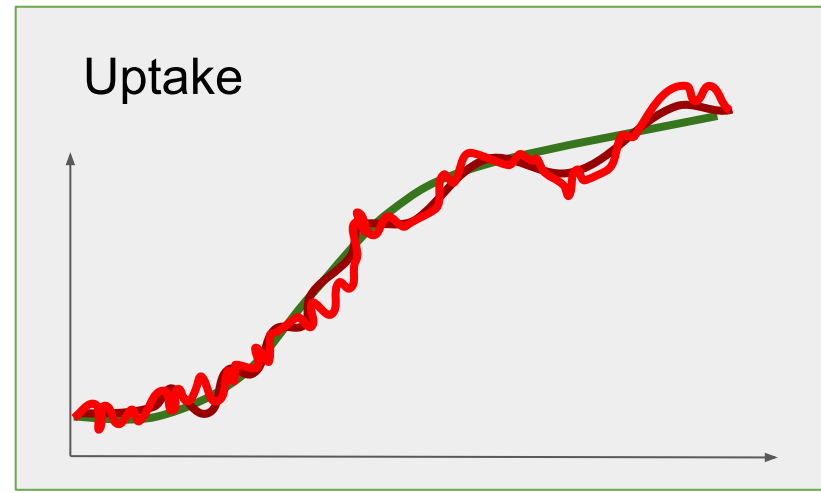
BMC Public Health

Conclusion

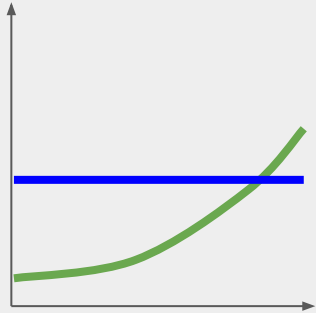
Overall, the review discovered positive attitudes towards the COVID-19 vaccine before February 2021, with 73% of the total survey participants reporting a high intention to receive the COVID-19 vaccine. COVID-19 vaccine acceptance can be influenced by many sociodemographic factors and individual risk perception towards COVID-19. The findings of this review imply that future research should explore the reasoning behind vaccine intentions for different sociodemographic groups, to allow targeted communication strategies to be formulated by governments and public health agencies. The impact of both vaccine availability and reported adverse effects must be monitored so public health policies can address these concerns. A high vaccine uptake to current mass-vaccination programmes and potential booster vaccinations is essential to achieve the end goal of herd immunity and combat any potential future variants.

Predicting Vaccine Uptake

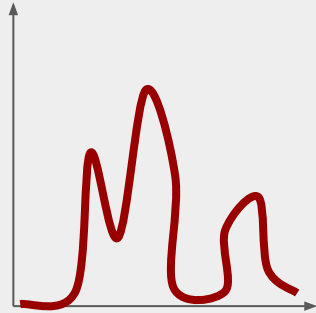
$$\text{Uptake}(t) = p * N * V_t / (p * N + V_t) + a * \text{mean}(C_{t-30 \dots t}) + b * \text{mean}(S_{t-7 \dots t})$$



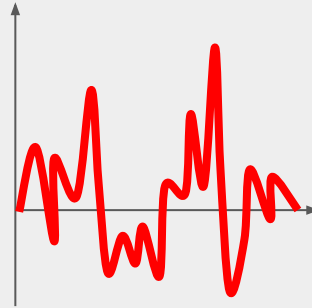
Vaccine availability V



Covid cases C

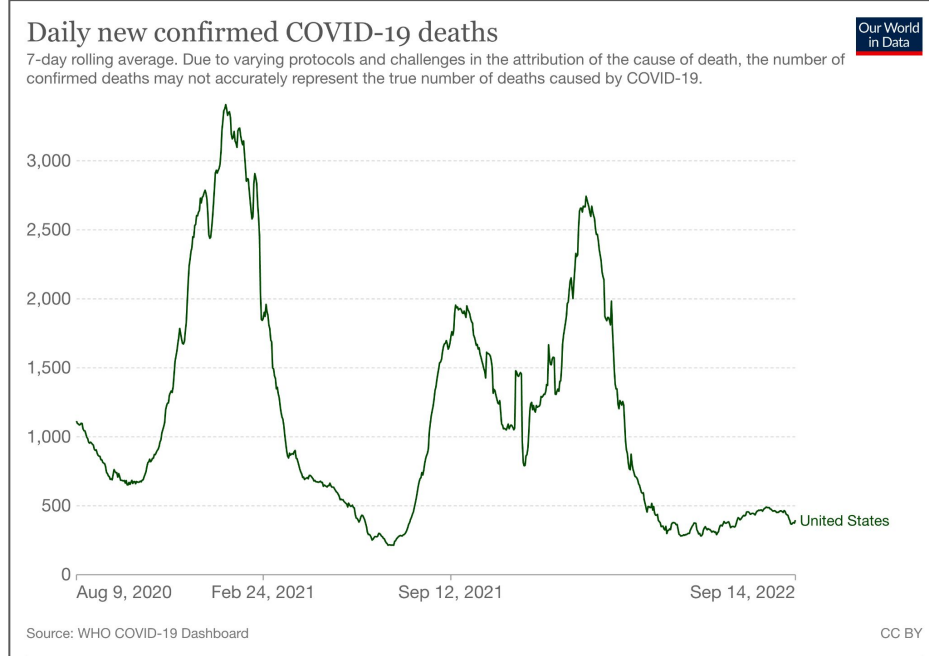
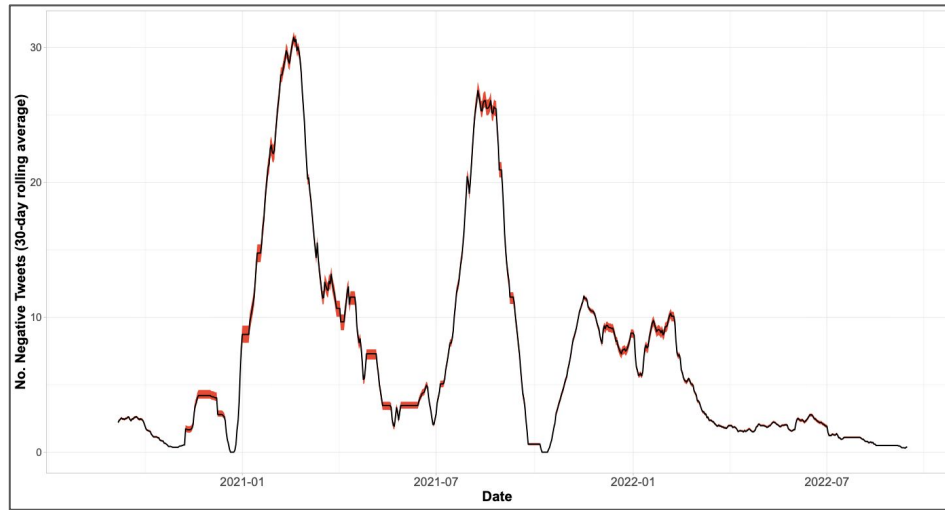


Vaccine sentiment S



Symbol	Meaning
N	Population size
p	Proportion of people willing to get vaccinated
V_t	Available vaccine doses at time t
C_t	Covid cases at time t
S_t	Sentiment at time t
a, b	Free parameters

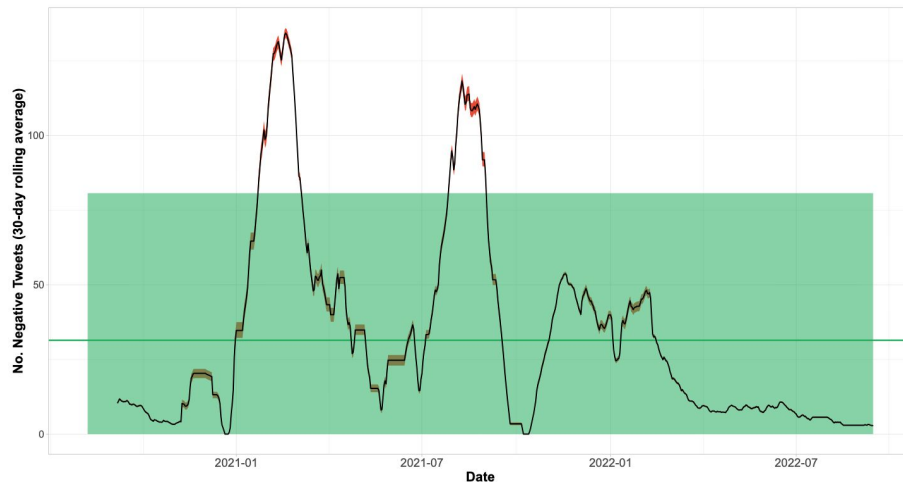
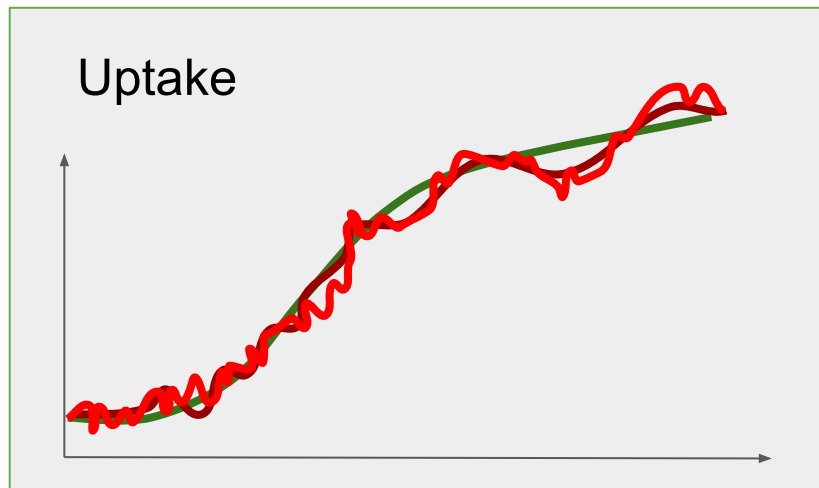
Vaccine sentiment analysis vs COVID deaths in the US



Questions

How does your proposed pipeline and platform handle predictions in real-time as well as longer time?

How would you incorporate uncertainty/ confidence of predictions in triggering alerts?



Threat Level Analysis

Threat Level Analysis Flowchart				
No of Negative Tweets			Threat Level	Proposed Intervention
Global	USA	UK		
> 80.7	> 17.9	> 6.8	RED	Take immediate action to counter negative sentiment
31.4-80.7	6.8-17.9	2.3-6.8	YELLOW	Take preventative measures to counter negative sentiment
< 31.4	< 6.8	< 2.3	GREEN	Continue monitoring

$> \text{mean} + \text{s.d.}$

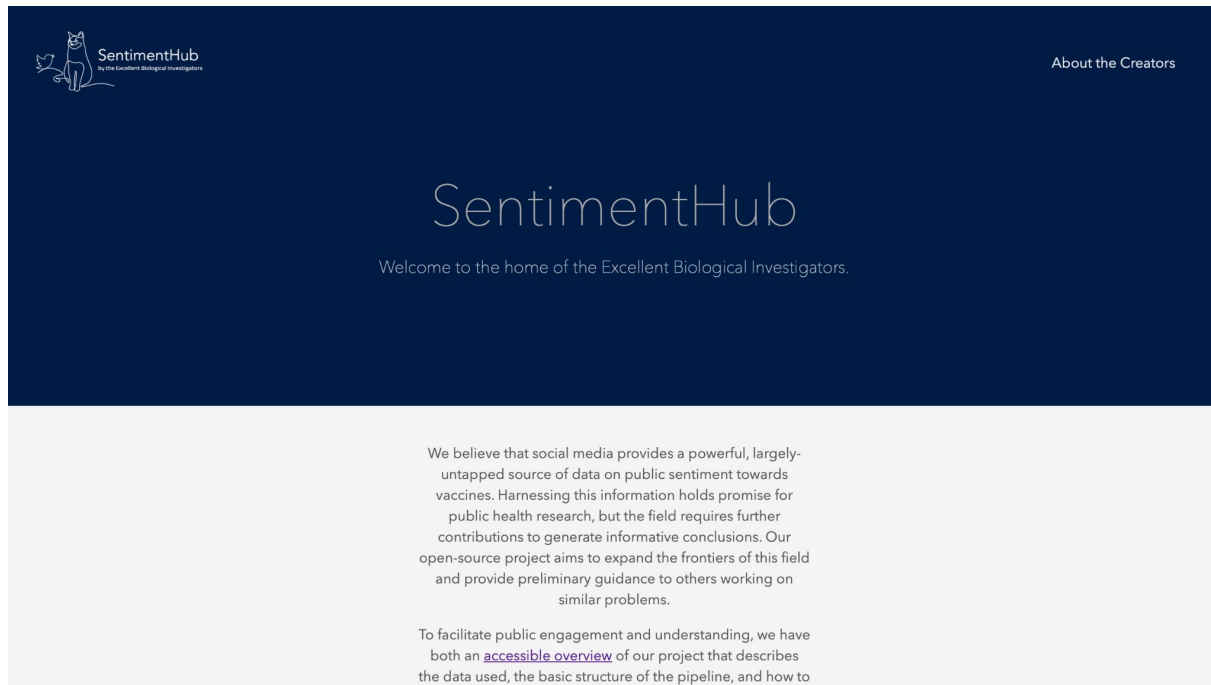
$\text{mean to mean} + \text{s.d.}$

$< \text{mean}$

How do we ensure public engagement & support?

- Familiarizing the public with how the methods work
- Data and privacy protection
- Use methods not interfering with user experience

Online interface: SentimentHub



Analysis for the public

Analysis (in-depth)

Quick access page

About us

Available at <https://qtoussaint.github.io/>

Conclusions

- SentimentHub uses open-source NLP to predict tweet sentiment and uncertainty, stratified by time and location.
- Twitter sentiment can be implemented into a model with the aim of more accurately predicting vaccine uptake.
- Sentiment analysis can be used where clinical surveillance has poor coverage.
- We have released SentimentHub as open-source code and as a public-facing website, with clear details on which data is collected and how it is used.

Future plans

- Considering tweet reach based on follower count.
- Relating negative sentiment to vaccine uptake in a cohesive mathematical model
- Predicting incidence from sentiment data
- Identifying misinformation using embedding
- Real-time data mining from social media sites
- Feedback button to collate concerns from public/research users

Bibliography

- Barbieri F., Camacho-Collados J., Espinosa-Anke L., Neves L. (2020), TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification, *Proceedings of Findings of EMNLP*
- Espinosa L., Wijermans A., Orchard F., Höhle M., Czernichow T., Coletti P., Hermans L., Faes C., Kissling E., Mollet T. (2022), Epiweetr: Early warning of public health threats using Twitter data. *Euro Surveill.*, ;27(39):2200177. doi: 10.2807/1560-7917.ES.2022.27.39.2200177
- Hayawi K., Shahriar S., Serhani M.A., Taleb I., Mathew S.S. (2022) ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public Health*. 2022 Feb; Volume 203, Pages: 23-30. doi: 10.1016/j.puhe.2021.11.022.
- Marcec R., Likic R. (2022) Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines, *Postgraduate Medical Journal*, Volume 98, Issue 1161, Pages 544–550, <https://doi.org/10.1136/postgradmedj-2021-140685>
- Qorib M., Oladunni T., Denis M., Ososanya E., Cota. P. (2023), Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset, *Expert Systems with Applications*, Volume 212, Page 118715, <https://doi.org/10.1016/j.eswa.2022.118715>
- Rakshit S. (2023), geograpy3, <https://pypi.org/project/geograpy3/>
- Salathé M., Khandelwal S (2011) Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. *PLOS Computational Biology* 7(10): e1002199.
- Terry, E., Cartledge, S., Damery, S. *et al.* Factors associated with COVID-19 vaccine intentions during the COVID-19 pandemic; a systematic review and meta-analysis of cross-sectional studies. *BMC Public Health* 22, 1667 (2022).