

A light gray world map is visible in the background of the slide, showing the continents of North America, South America, Europe, Africa, Asia, and Australia.

Tracking health trends using automated real-time social media text classification

Martin Müller

 @marmuel_

Digital Epidemiology Lab, EPFL

Sentiment analysis: A short story



You!



Sentiment analysis



Heaps of social
media text data

The answer to your
research question

Approach #1:

Goal: Understand what people think about vaccines

Method:

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
analyzer = SentimentIntensityAnalyzer()
for sentence in sentences:
    sentiment = analyzer.polarity_scores(sentence)
```

Result:

"Great to see the ban of vaccines"

positive

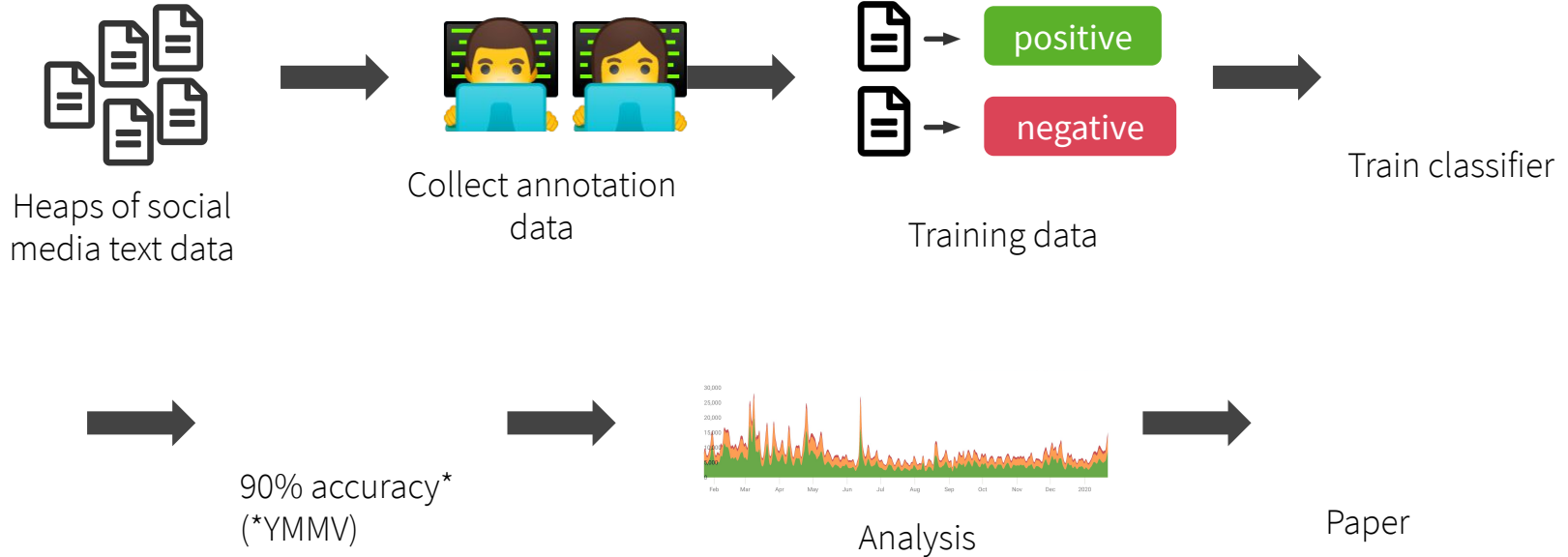
"We need to stop anti-vaccine groups online"

negative

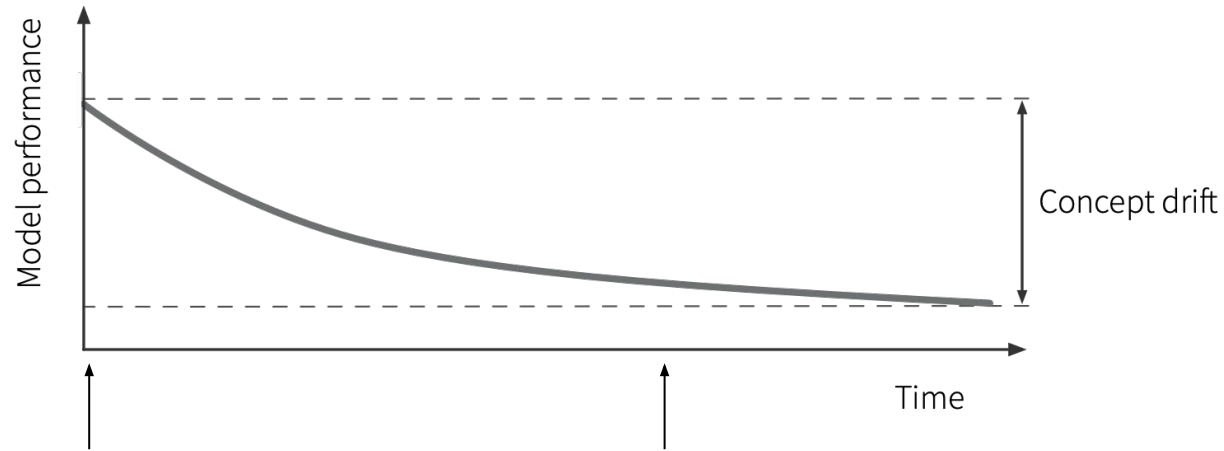
Conclusion approach #1:

We want **stance** not **sentiment**

Approach #2: Supervised Machine Learning



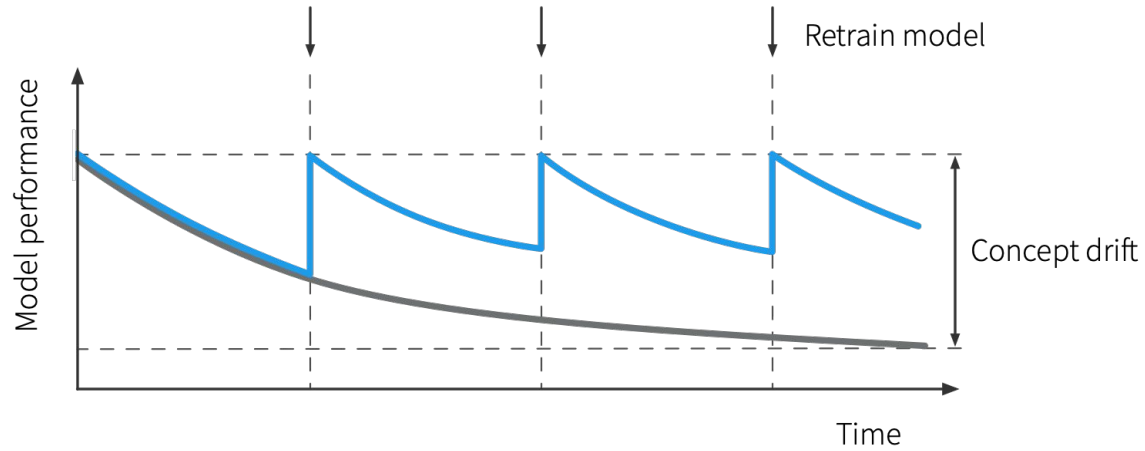
Enter: Concept drift



Your paper

You are here

Combatting concept drift

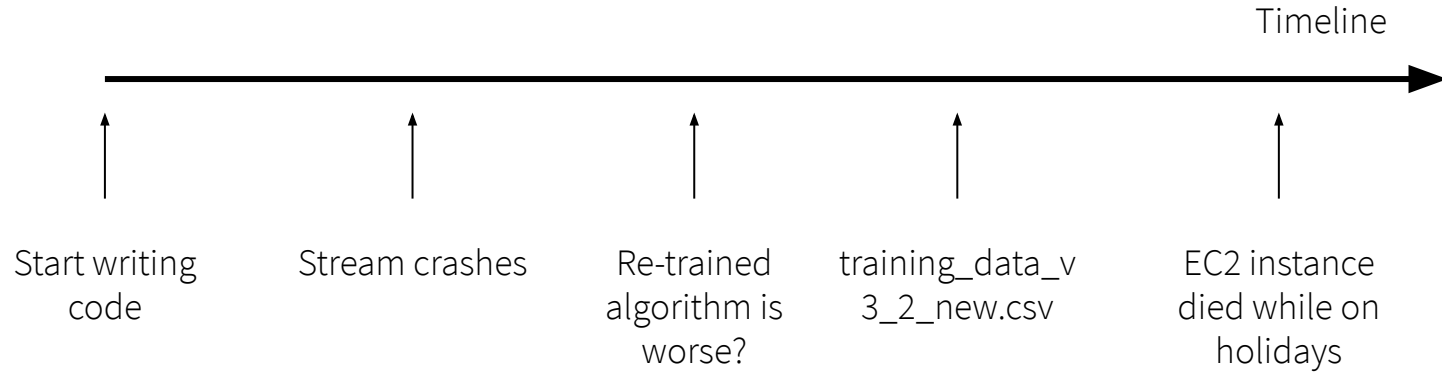


Conclusion approach #2:

We need **continuous**

- annotation
- retraining
- analysis

Approach #3: Manually retraining



Conclusion approach #3:

- Avoid manual intervention
- Monitor everything
- Infrastructure is key

Help us track health trends in real time.

166,577 tweets labelled so far

Contribute

Find out more



Set Apart Ones
@setapartones



We barely got into the new decade, already there's a coronavirus scare. Australia fires & weather calamities etc... Strapped up guys this decade is surely going to be a rough one, my spirit grieves 4 those still rejecting the Messiah. #wuhanvirus

8:15 AM - Jan 23, 2020



[See Set Apart Ones's other Tweets](#)



annotated 7 hours ago · by guest
in project Novel Coronavirus (nCoV) - Wuhan disease outbreak

Crowdbreaks



Key technologies:



Flask

crowdbreaks-streamer



crowdbreaks



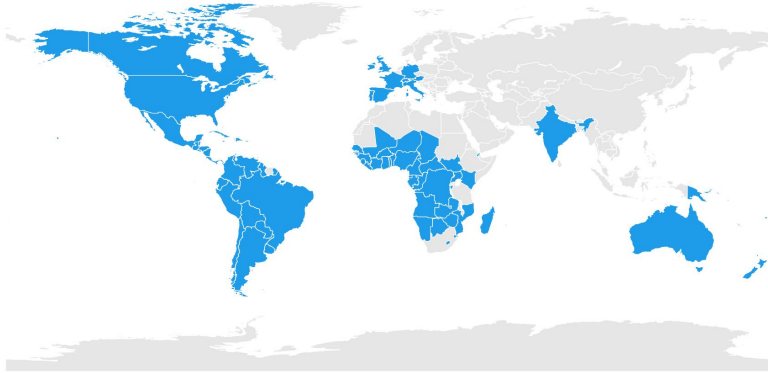
AWS Sagemaker



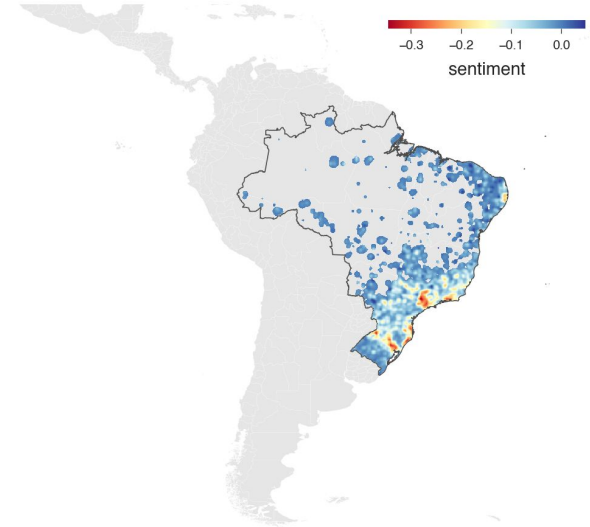
 <https://github.com/crowdbreaks> (fully under MIT license)

Current projects

Vaccine confidence



Currently in 6 languages (135 countries)



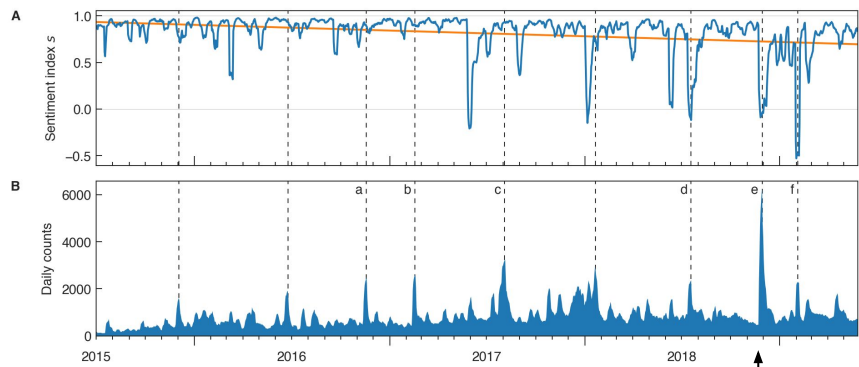
Challenge:

Applying cross-lingual transformer models for non-English languages

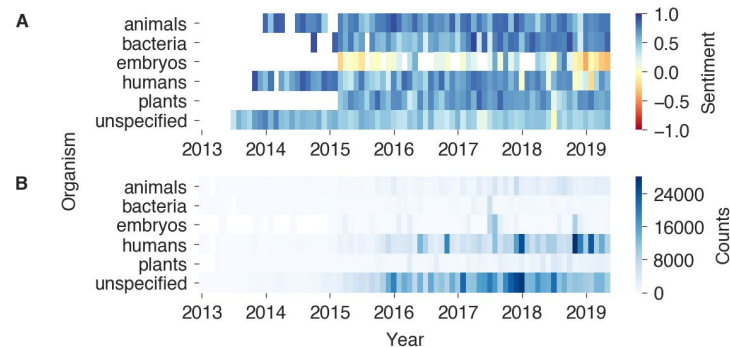
Collaboration with



CRISPR sentiment



"CRISPR babies"
scandal



BERT text classification



Relevance
 $F_1=0.91$

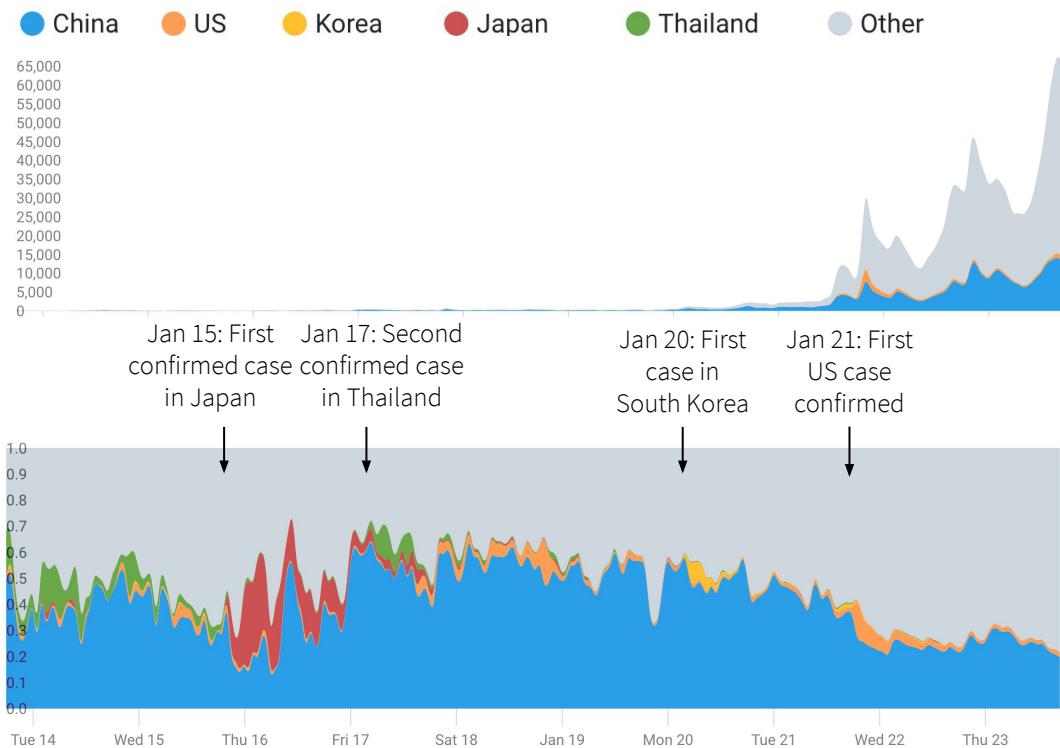


Sentiment
 $F_1=0.73$



Organism
 $F_1=0.89$


Novel Coronavirus outbreak (nCov)



- Information flow is very high!
- Lots of misinformation which is picked up by news outlets

⇒ Annotation, training and analysis needs to happen within days!

 Main work in progress: 

- Cross-lingual text classification
- Continuous AI benchmarks
- Real-time visualizations 

Thanks for your attention!

We're looking for PostDocs! (& PhDs!)