

# Automatic Identification of 5C Vaccine Behaviour on Social Media

Ajay Hemanth, **Aminath Shausan**, Gianluca Demartini, Afshin Rahimi



@W-NUT 2022 (Collocated with COLING 2022), Rep Korea

October 16, 2022



Safe Blues

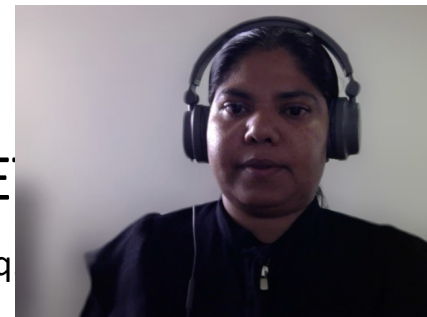
<https://safeblues.org>

AI4PAN

<https://ai4pandemics.org>

ID-NE

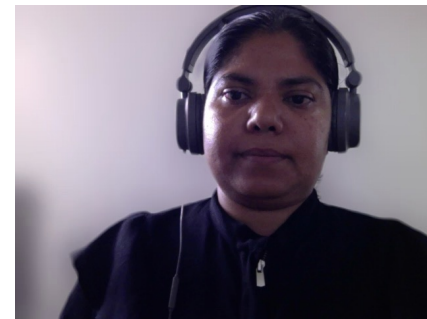
<https://public-health.uq>



# The 5C Model

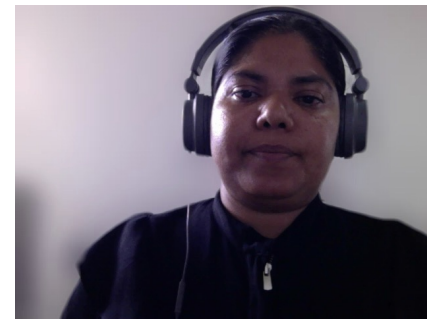
A theoretical framework for assessing an individual's psychological reasoning towards vaccination (Betsch et al. 2018).

Behaviour	Description
Confidence	Trust in the safety and effectiveness of vaccines and the health system
Constraints	Structural and psychological barriers
Complacency	Not perceiving diseases as high risk
Calculation	Engagement in extensive information searching
Collective Responsibility	Willingness to protect others



# Aim

To investigate Australian public attitude towards COVID-19 vaccine using Twitter data and the 5C model



# Data Collection & Annotation

## Data collection

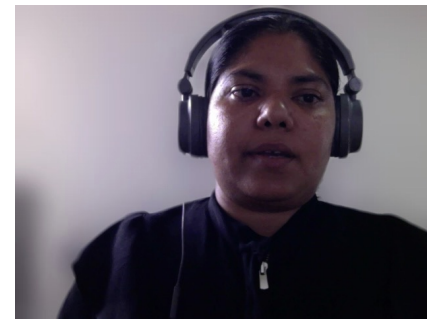
- Collected 60,000 tweets related to COVID-19, restricted to Australia
- Time frame: *01 Feb 2020 – 30 Oct 2021*
- keywords: *vax, vaccine, vaccinate, vaccination, jab, Pfizer, astrazeneca*

## Annotation

A weighted sample of 20,000 tweets.

Step 1: sentiment annotation

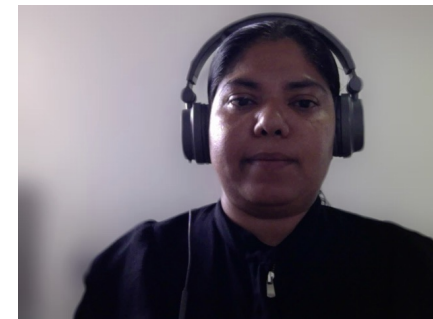
Step 2: 5C annotation



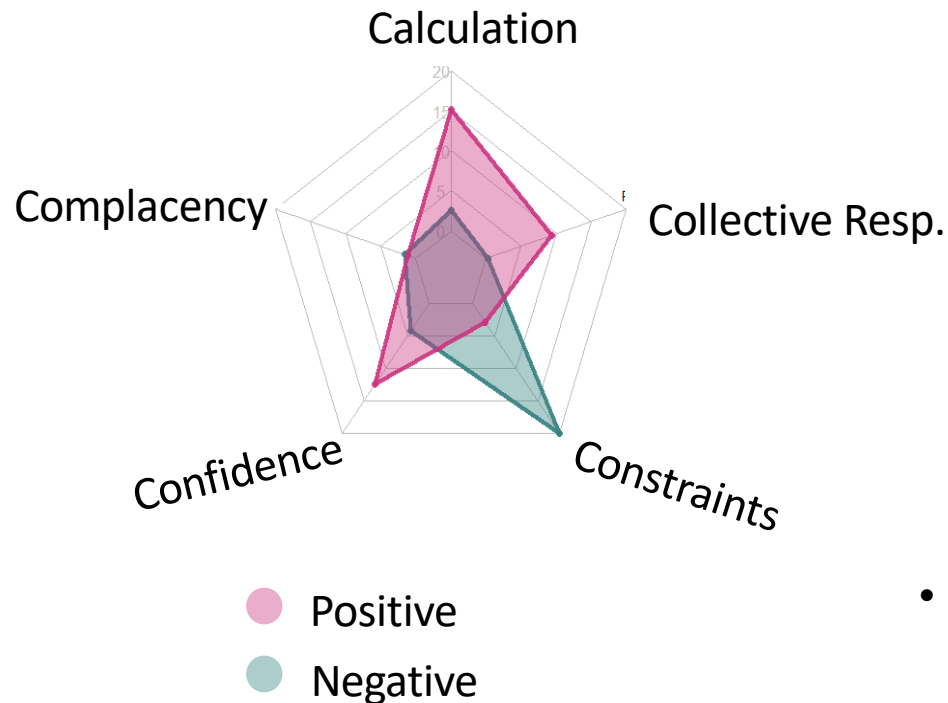
# Annotated Dataset:- A Sample

Label	Sample Tweet (rephrased for privacy reasons)
+ Complacency	It takes just one COVID infected person to start a Pandemic. Please vaccinate.
+ Collective Resp.	Difficult to wear mask but we have to. Vaccinate, not just for yourself but for others too.
+ Calculation	Vaccines don't prevent infections, they prevent severe disease and hospitalisation.
+ Constraints	no jab no job. That's how it should be.
- Complacency	99% of people recover from COVID, vaccines shouldn't be mandated.
- Confidence	A vaccine that might kill you for various reasons while big pharma benefits.
- Calculation	myocarditis risks is higher in teenage boys, we shouldn't rush into vaccinating them.
- Constraint	AZ age limit should be 60, not 50

**Note:** Original tweets rephrased for privacy reasons

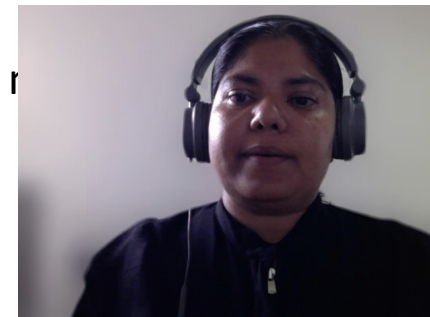


# Distribution of Sentiments and 5C Categories

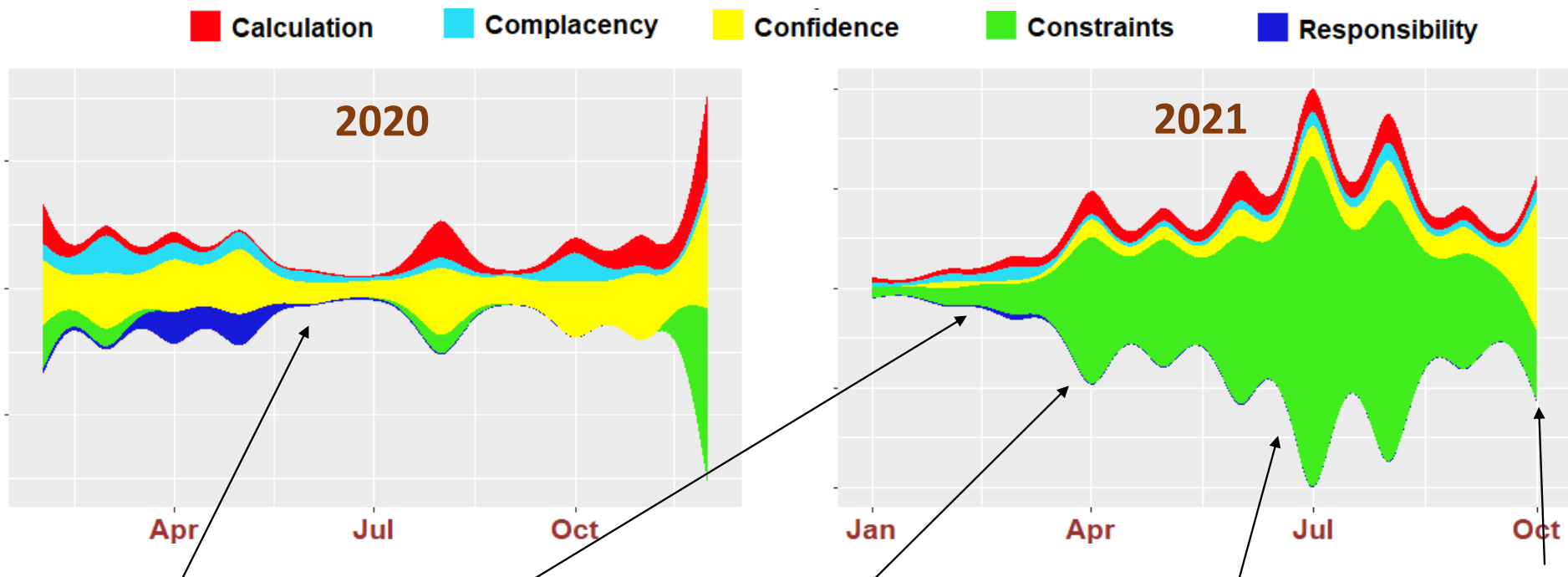


Label	Bigrams
Constraints	vaccine rollout, scott morrison, aged care
Confidence	az vaccine, catching measles, pfizer vaccine
Complacency	COVID19 vaccine, blood clots, cold flu
Calculation	blood clots, sore arm, az vaccine
Collective Resp.	aged care, stay safe, vaccine hub
Positive	az vaccine, pfizer vaccine, sore arm
Negative	vaccine rollout, az vaccine, scott morrison

- Negative emotions -> vaccine rollout, AZ vaccine and the government
- Positive emotions -> vaccine types and effects



# Evolution of the Negative 5C Categories



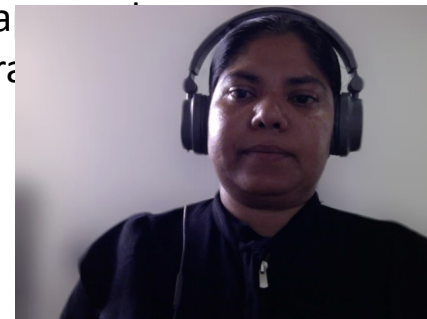
A high approval rating for govt. due to proactively closing border, lockdown opening up super funds, securing citizens

Ignoring Pfizer offer for vaccine import backfires for govt.

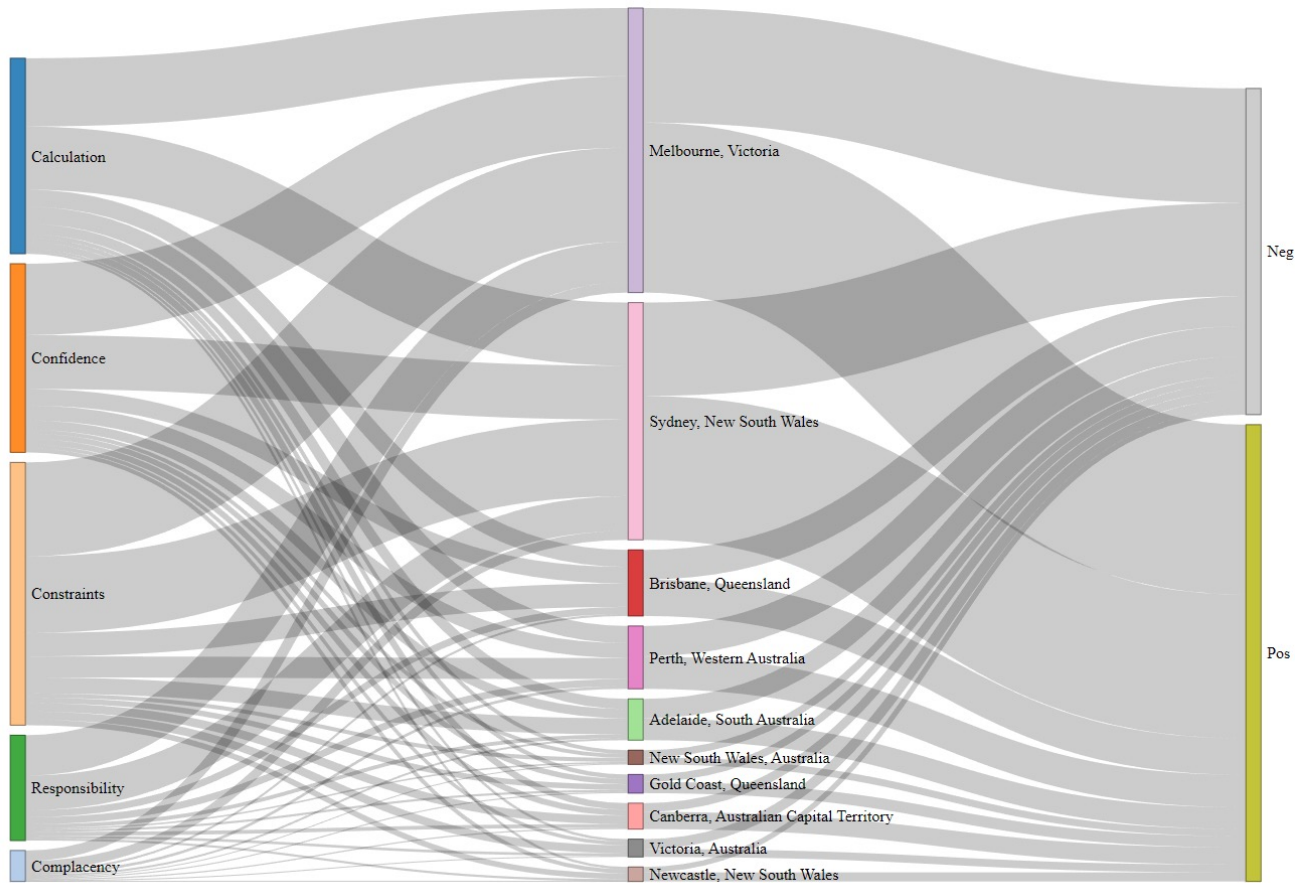
AstraZeneca, a locally manufactured vaccine causes panic amid blood clot claims.

Lack of vaccine combined with AZ blood clot fear and ignoring Pfizer offer causes public outrage

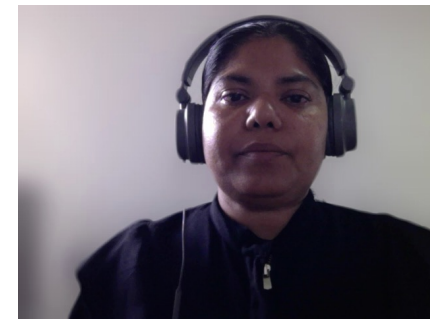
Govt.  
a  
ra



# Location Based Inference



- Majority of tweets are from Sydney and Melbourne
- Both cities show identical behaviour with respect to the 5C scale
- Negative attitudes -> Confidence and Calculation
- Positive attitudes -> Constraints, Collective Responsibility, Confidence



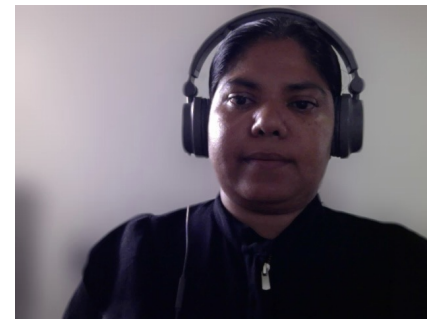


# Vaccine Behavior Identification

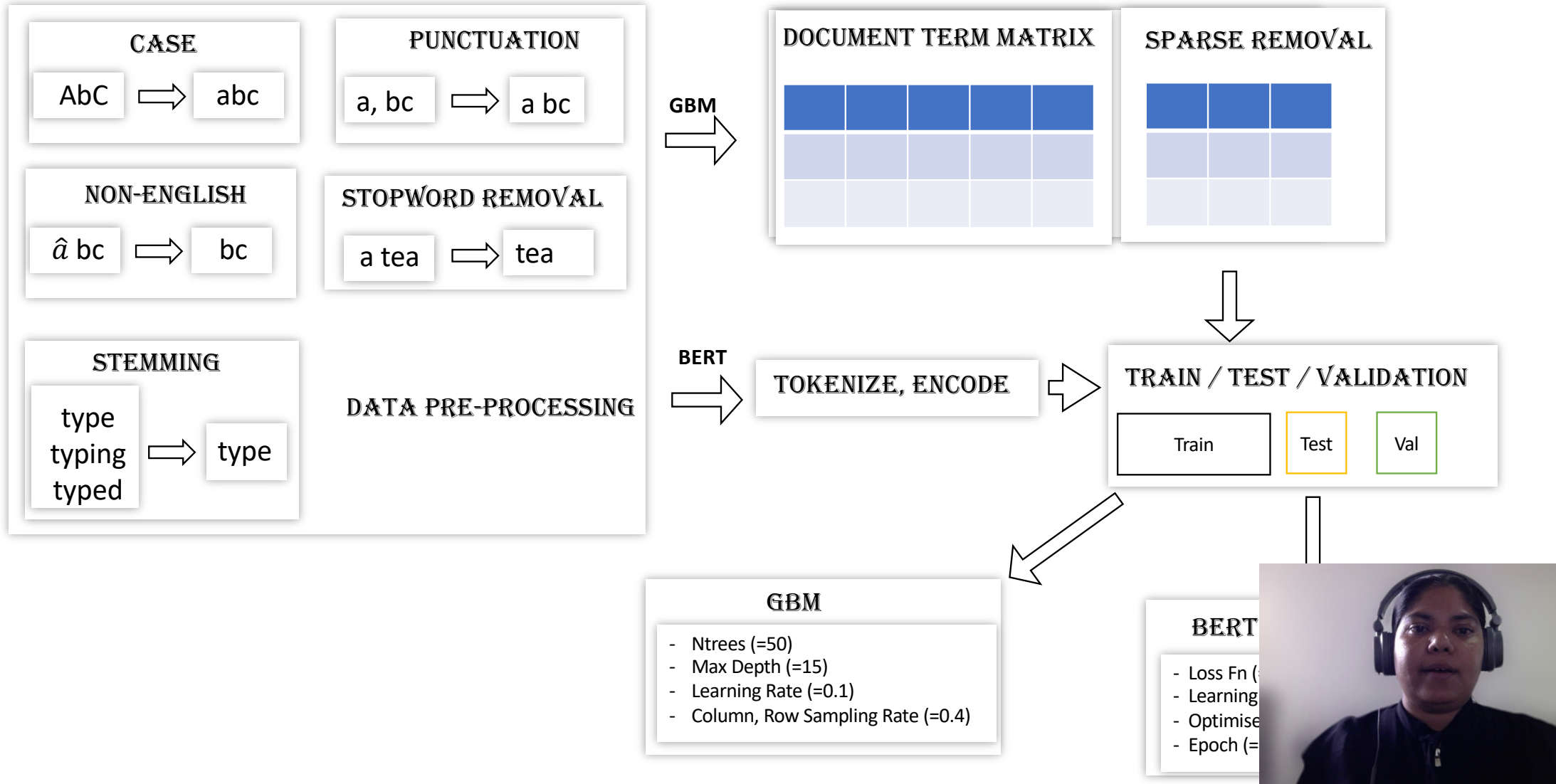
**Task 1: Sentiment identification** -> binary classification problem

**Task 2: 5C categorisation** -> multi-class classification problem

**Task 3: Combined sentiment & 5C categorisation** -> multi-class classification problem

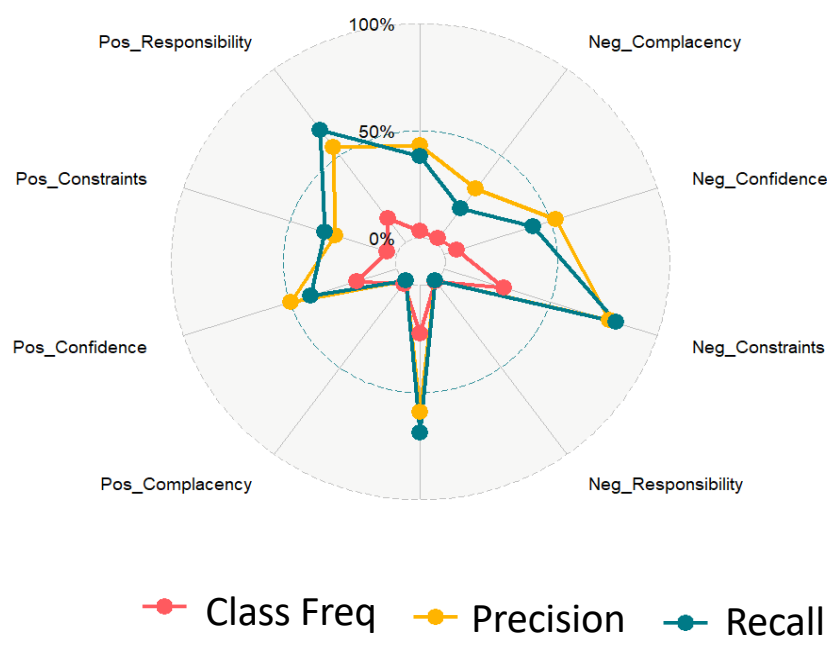


# Modeling Process

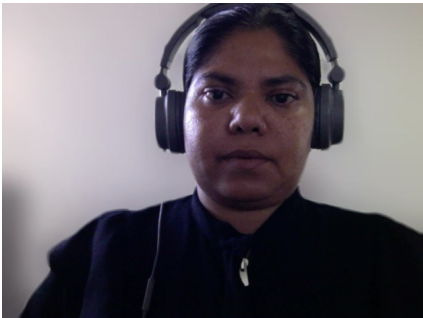
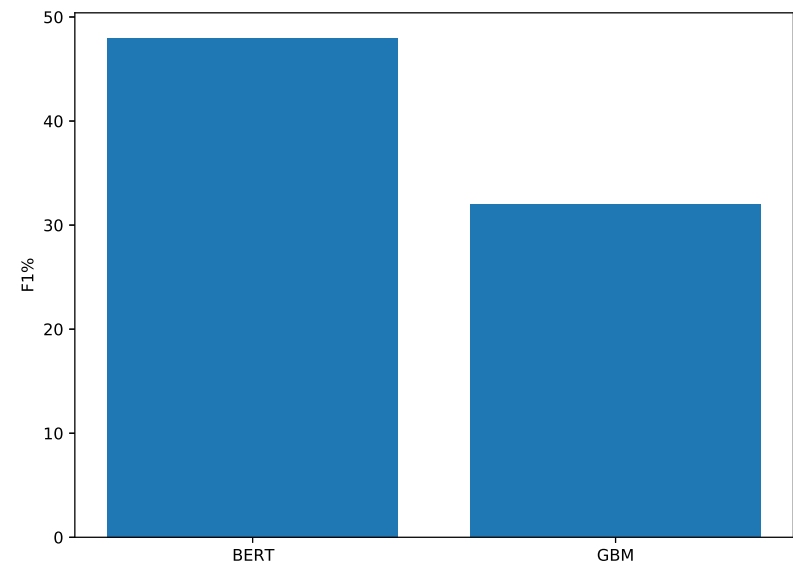


# Results

Performance of BERT in Tasks 3 with class distribution

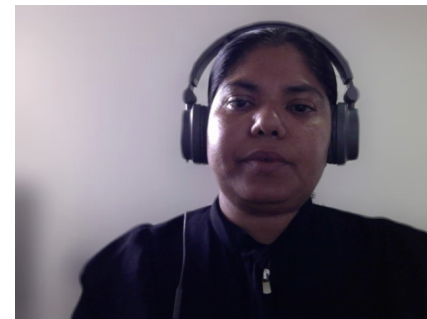


Comparison of BERT and GBM in Task 3 prediction



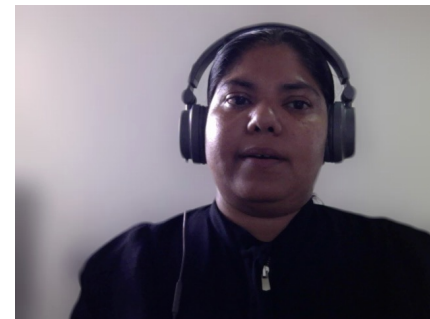
# Contributions

- First large-scale vaccine behaviour Twitter dataset
- Two benchmark models
- Highlight challenges due to imbalanced data



# Limitations, Challenges & Extensions

- Some tweets -> no clear positive or negative stance -> ignored
- Some discretions for labelling positive 5C categories  
Eg; tweets which mentioned only about taking vaccine without any further -> positive Confidence
- GBM and BERT are affected by class imbalance -> increasing weights for low frequency classes reduced accuracy
- Extension -> other social media domains (Eg; Facebook and Google Trends)



# Acknowledgements & Contact

**Funding** from UQ AI4PAN group

**Data:** <https://github.com/ajayhemanth/5C-Twitter>

## Contact



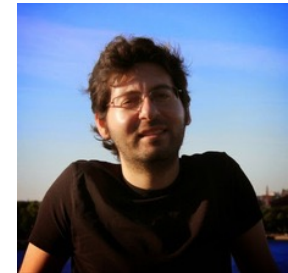
Ajay Hemanth  
[a.sampathkumar@uq.edu.au](mailto:a.sampathkumar@uq.edu.au)



Aminath Shausan  
[a.shausan@uq.edu.au](mailto:a.shausan@uq.edu.au)



Gianluca Demartini  
[g.demartini@uq.edu.au](mailto:g.demartini@uq.edu.au)



Afshin (Ash) Rahimi  
[a.rahimi@uq.edu.au](mailto:a.rahimi@uq.edu.au)



Thank you!