Automatic Identification of 5C Vaccine Behaviour on Social Media

Ajay Hemanth Sampath Kumar **Aminath Shausan**
Gianluca Demartini **Afshin Rahimi**
The University of Queensland
{a.sampathkumar,a.shausan,g.demartini,a.rahimi}@uq.edu.au

Abstract

Monitoring vaccine behaviour through social media can guide health policy. We present a new dataset of 9471 tweets posted in Australia from 2020 to 2022, annotated with sentiment toward vaccines and also 5C, the five types of behaviour toward vaccines, a scheme commonly used in health psychology literature. We benchmark our dataset using BERT and Gradient Boosting Machine and show that jointly training both sentiment and 5C tasks (F1=48) outperforms individual training (F1=39) in this highly imbalanced data. Our sentiment analysis indicates close correlation between the sentiments and prominent events during the pandemic. We hope that our dataset and benchmark models will inform further work in online monitoring of vaccine behaviour. The dataset and benchmark methods are accessible online.[1]

1 Introduction

The development of effective and safe vaccines has been shown as one of the most successful means to mitigate the spread of COVID-19 disease around the globe. As with any other vaccination program, COVID-19 intervention in any country depends on public acceptance and vaccine uptake. This has to be done by a large proportion of society to attain herd immunity, which is estimated to range from 67% to 95% (Mills et al., 2020; Randolph and Barreiro, 2020; Anderson et al., 2020). However, vaccine hesitancy has been identified as one of the top 10 challenges to global health by the World Health Organization in 2019 (World Health Organization, 2019 January [cited 14 August 2022]).

Vaccine hesitancy refers to the delay in acceptance or refusal of vaccination by the public despite the availability of vaccination services (MacDonald et al., 2015). This behaviour has been significantly prominent towards COVID-19 vaccines as they differ from previous vaccines in many respects: accelerated development, novel techniques used, potential side effects, uncertainty regarding the size and extent of their effectiveness, and limited production compared to the demand (Dubé and MacDonald, 2020).

Australia began its mass vaccination program in late February 2021 but was initially hampered by low vaccine uptake due to mistrust of vaccine effectiveness and the government, the blood clotting syndrome (TTS) associated with the AstraZeneca product, shipment delays, and misinformation (Kaufman et al., 2022b). Despite the fact that the present uptake is high ($> 90\%$ two doses for over 16 years), the coverage varies across age, jurisdictions, and the number of doses (Australian Government: Department of Health and Aged Care, 2022 August [cited 14 August 2022]). High coverage is primarily driven by travel desires (and travel vaccination requirements) and the need to mitigate risk. As new variants of SARS-CoV-2 are emerging and existing immunity waning in a short period of time, it is likely that people will need to get multiple booster doses. Consequently, identifying immediate and future public behaviour toward vaccination is important for public health authorities to combat possible challenges to reaching and maintaining herd immunity.

The 5C model provides five measures (Confidence, Complacency, Constraints, Calculation and Collective Responsibility) for assessing an individual's psychological reasoning towards vaccination (Betsch et al., 2018, 2020). *Confidence* associates the trust in vaccine effectiveness, safety, and the system that delivers it. People with low confidence mistrust the healthcare system, fall for misinformation, believe in conspiracies, and doubt the benefit of vaccines. *Complacency* exists when vaccination is viewed as a low priority or when vaccine-preventable diseases are not of a concern.

[1] https://github.com/ajayhemanth/5C-Twitter

Proceedings of the 2022 COLING Workshop: The 8th Workshop on Noisy User-generated Text (W-NUT 2022), pages 136–146

High complacency correlates with low uptake of a vaccine. *Constraints* refer to the structural and psychological barriers to getting vaccinated. For example, geographical constraints in accessibility, limited language, and health literacy, and cost may postulate high constraints. *Calculation* defines the engagement in extensive information searching, which may lead to lower vaccination willingness arising with exposure to a high volume of anti-vaccination content. *Collective responsibility* asserts an individual's willingness to protect others by getting vaccinated and contributes to herd immunity. See Table 1 for a summary of the description of the 5C model.

Most works in identifying vaccine behaviour use surveys that are costly, time-consuming, and don't scale to a large population. The use of social media for expressing vaccine-related opinion provides a great opportunity to use online social monitoring tools to guide health policy-making for vaccine adoption. There are few recent studies that adapt the 5C scheme for online health monitoring. Greyling and Rossouw (2022) focus on the analysis of tweets rather than building predictive models, Fues Wahl et al. (2022)'s uses 1794 tweets from Scandinavian users and manually categorise them into the 5C categories, and Boucher et al. (2021) focuses on the vaccine trials and uses unsupervised methods.

Our contributions are as follows: 1) we provide the first large-scale vaccine behaviour dataset in English, annotated with both 5c and sentiment; 2) we present two benchmark models and show the challenges of 5C predictive models given the highly imbalanced data; and 3) we analyse the data showing that changes in 5C distribution is not uniform across regions, indicating opportunities for targeted health messaging. We make the dataset and benchmarks available hoping to impact future work in online vaccine behaviour monitoring based on the 5C framework.

2 Related work

The 5C model has been widely applied to examine COVID-19 vaccination behaviour. In 2020, Kwok et al. (2021a) estimated Hong Kong nurses' intention to receive COVID-19 vaccine using the 5C model and examines the correlation of their vaccine behaviour to previous influenza vaccination. Thunström et al. (2021) applied the 5C scale to investigate psychological reasoning behind the previous uptake of measles and flu vaccines by adults in the United States and their intention to get COVID-19 vaccination for themselves and their children. Wismans et al. (2021) studied the psychological drivers of vaccination intention in university students across the Netherlands, Belgium and Portugal, using the 5C model. Gallant et al. (2021) implemented the 5C model to investigate older adults' vaccination behaviour in the United Kingdom over the first year of the pandemic. Lindholt et al. (2021) examined the levels and predictors of acceptance of an approved COVID-19 vaccine in eight Western countries by utilizing the 5C model. Rustagi et al. (2022) applied the 5C model to identify vaccine hesitancy among chronic disease patients availing care in a primary health facility in India.

Previous studies examined COVID-19 vaccination behavior utilizing traditional surveys (Seale et al., 2021; Trent et al., 2022; Rhodes et al., 2021; Edwards et al., 2021; Dodd et al., 2021; Kaufman et al., 2022a; Kwok et al., 2021a; Thunström et al., 2021; Sherman et al., 2021; Wismans et al., 2021; Paul et al., 2021; Sallam, 2021; Akarsu et al., 2021; Fisher et al., 2020; Freeman et al., 2020; Ward et al., 2020; Lazarus et al., 2021). However, such surveys are often costly in their design and implementation, time-consuming, produces limited data and represent comparatively short-term situation. Recently, Twitter has been increasingly applied in research concerning public attitude towards vaccination due to its advantages of availability of a large amount of real-time posts without any costs, ease of access and public searching facility. In spite of these advantages, there remains a gap in Australia for Twitter-based COVID-19 vaccination research. We found just one study (Kwok et al., 2021b), conducted during 2020, which addresses Australian public opinion towards COVID-19 vaccine solely from Australian Twitter users. Sentiment analysis of this study has shown that the majority of people expressed positive emotions towards vaccine with trust and anticipation as the most prominent behaviors associated with it, while fear being the top negative emotion.

An investigation of Twitter posts from 10 countries, including Australia, has revealed that more information about vaccines' safety and the expected side effects may increase public positive attitudes towards vaccination Greyling and Rossouw (2022). Similarly, sentiment analysis from 4 million tweets

Label	Behavior	Description
C1	Confidence	Trust in safety and effectiveness of vaccines and health system
C2	Constraints	Structural and psychological barriers
C3	Complacency	Not perceiving diseases as high risk
C4	Calculation	Engagement in extensive information searching
C5	Collective responsibility	Willingness to protect others

Table 1: Description of 5C categories. Collective Responsibility and Responsibility (Resp.) are used interchangeably in this work.

across several nations, including Australia, has found the prevalence of vaccine hesitancy and objections outweighs vaccine interests (Yousefinaghani et al., 2021). Various other studies have assessed COVID-19 vaccination behavior using Twitter data from a specific country or a particular region. These include, studies based on posts from the United States (Jang et al., 2021; Engel-Rebitzer et al., 2021; Germani and Biller-Andorno, 2021), the United Kingdom (Hussain et al., 2021), Japan (Niu et al., 2022b,a), China (Gao et al., 2021; Wang et al., 2020), Canada (Griffith et al., 2021), Africa (Gbashi et al., 2021). Other studies address COVID-19 vaccination behaviour at a global scale (Chopra et al., 2021; Lyu et al., 2021; Xue et al., 2020).

Along with sentiment analysis, some studies have implemented the 5C model to examine COVID-19 vaccination emotions as well as attitudes towards other vaccines in tweets. Greyling and Rossouw (2022) constructed a multiple linear regression model to examine positive attitudes towards vaccines across ten countries. In their model, among other methods, the positive 5C categories were applied to identify the covariates. Boucher et al. (2021) applied a topic modeling approach to investigate the mistrust in Covid-19 vaccination based on English and French tweets. They then used the 5C scale to categorize the topics and found that all mistrusts fell into the Confidence category. Fues Wahl et al. (2022) applied the 5C model to map relevant predictors for several vaccination behaviours in Scandinavian Twitter users. Similar to our work, they manually labelled each tweet according to the 5C scale. However, unlike our dataset, their dataset was multi-labeled, as each tweet was assumed to have multiple 5C categories. They did not, however, include Covid-19 vaccination as a specific vaccination category in their analysis.

3 Method

3.1 Dataset

We made use of the Twitter API to collect $60,000$ COVID-19 vaccine-related tweets restricted to Australia from 01 February 2020 to 30 October 2021, using keywords: *vax, vaccine, vaccinate, vaccination, jab, pfizer, astrazeneca*. We selected a weighted sample of $20,000$ tweets with weights proportional to the total number of tweets within the time frame of a week and annotated each tweet first based on its sentiment towards vaccination and then its associated 5C category if the tweet contains either a positive or negative stance. We note that some tweets do not provide a clear positive or negative stance and thus have been categorized accordingly if the stance is totally irrelevant to the topic of vaccination, or if the stance concerns vaccination, but does not belong to either the positive or negative category. We excluded tweets from the last two categories in our analysis. We labeled each tweet based on the most prominent 5C behaviour. Thus, our dataset has the structure of binary categories in terms of tweet's sentiment and multi-class categories with regard to 5C vaccine behaviour. Table 2 depicts a sample of labeled tweets. The final dataset consists of 9471 annotated tweets. The study has received clearance by the authors' organisation's human research ethics committee.

We note that, in the original 5C model (Betsch et al., 2018) much of the description was made for the negative aspects of the model, hence we made some discretion for labeling positive 5C categories, as, based on the tweet texts, these categories were not clearly identified. For example, tweets which mentioned only about taking vaccine without any further explanations (like 'got my first vaccine !!!') were labeled as positive Confidence. Tweets which talked about getting vaccine as a compulsory action (like 'no jab no job') were labeled as positive Constraints. Similarly, tweets which supported vac-

Label	Sample Tweet (rephrased for privacy reasons)
+ Complacency	It takes just one COVID infected person to start a Pandemic. Please vaccinate.
+ Complacency 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

Table 2: A sample of eight tweets (rephrased for privacy reasons) and their labels. + and - refer to the positive and negative sentiments towards COVID-19 vaccination, respectively. The five C categories are also shown (see Table 1 for definition).

cine based on data as well as personal experience (like 'I took AZ vaccine 1 week ago, and there is no symptoms of blood clot') were categorized as positive Calculations. To check the validity of the annotations, we compared annotation agreements between two researchers from the team who independently labeled a random sample of 200 tweets. Using the Cohen's kappa (Cohen, 1960; McHugh, 2012) statistic, we found a strong ($\kappa = 0.95$) level of agreement between the researchers with regards to the sentiment labels (see Table 3 for contingency table), and a strong ($\kappa = 0.88$) agreement for the 5C labels (see Table 4 for contingency table). All 200 tweets have been used in the sentiment label comparison while only those tweets which both annotators agreed as positive and negative have been included in the 5C label comparison.

A/B	InSuff	Neg	Pos	X
Insuff	19	0	2	0
Neg	0	68	0	0
Pos	1	0	82	1
X	0	2	1	24

Table 3: Summary of annotation agreements between two researchers for a random sample of 200 tweets. Labels indicate positive (Pos), negative (Neg), irrelevant (Insuff) and inconclusive (X) stance towards vaccination.

3.2 Data Analysis

Overall, people expressed significantly high negative emotions in the Constraints category, similar levels of positive and negative emotions in the Complacency category and have been otherwise always positive towards vaccination depicting high Confidence and Calculation behaviours (Figure 1). The most eminent concerns regarding negative emotions have been related to the constraints caused by vaccine roll out, Pfizer vaccine, aged care facilities, and the government and its leader (Table 5). On the other hand AstraZeneca and Pfizer vaccines and COVID vaccine in general lead among positive topics, with high calculations and confidence associated with them (Table 5). A high number of tweets with the Positive Calculation category relate to the tweets in support of AstraZeneca with their own experience as evidence to disprove the tweets associating blood clot issues with that vaccine. This causes AstraZeneca to be a dominant unigram in both positive and negative instances.

A/B	C1	C2	C3	C4	C5
C1	75	0	2	1	1
C2	0	27	1	1	2
C3	0	0	4	0	1
C4	0	0	0	4	0
C5	0	1	0	0	17

Table 4: Annotation agreement table for 5C (See Table1 for the definition of C1-C5).

As the pandemic prevailed from 2020 to 2022, negative attitudes towards vaccination continuously varied in terms of prevalence and corresponding 5C reasoning. During 2020, the prevalence of negative stances has been relatively stable apart from the peaks around August and December, and the

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

Table 5: Top 3 most frequent bigrams categorized by their sentiment and 5C classes.

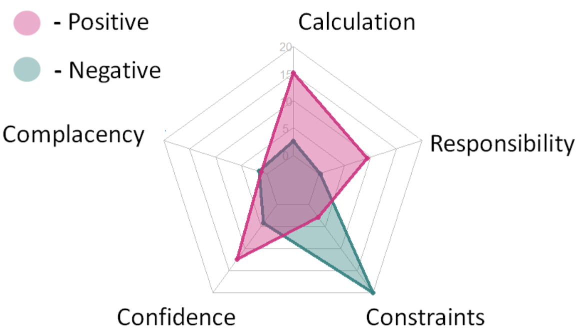

Figure 1: Distribution of positive and negative 5C.

shallow dips between the end of May to the end of July and in September (Figure 2 (left)). The shallow period from the end of May 2020 to the end of July 2020 coincides with the event of high approval towards the government for proactively closing the border, implementing lockdown, releasing super funds. Negative stances have been stable from early 2021 until around the end of March, and peaked during April, possibly due to the panic caused by the blood clot from AstraZeneca vaccine. Multiple peaks have then appeared with prominent ones occurring in July, August and December 2021 (Figure 2 (right)). The peaks in July and August correlate with the public outrage caused by the shortage of vaccine doses, the continuous emergence of blood clotting incidences from AstraZeneca vaccine, and the government ignoring to use the offer of Pfizer vaccine from that company. Consequently, the peak in December 2021 correlates to high displeasure towards the government at that time.

From the beginning of 2020 until around November that year, people voiced strong lack of confidence with regard to vaccination, but has then changed to the constraints from December 2021 onward until September 2021 where the notion has turned back to lack of confidence. Collective responsibility has not been consequential much to negative emotions except during April 2020 to March 2020 (Figure 2 (left)). This may be due to the fact that people resisting to lockdowns around April 2020, pointing their right to freedom. Such emotions have reduced after June 2020, with many deaths due to Covid occurring during April 2020. High level of negative confidence seen from October 2020 to the end of that year may be related to the events of the development of vaccines and people being doubtful of their effectiveness due to some vaccinated people in other countries getting re-infected. The emergence of negative Complacency and Collective responsibility from February 2021 to March 2021 may happen because of public resistance to the government's allocated vaccine, pointing out their right to choose the vaccine type. High level of negative Constraints emotion depicted from April 2021 onward may be a consequence of combination of several factors: shortage of vaccine, fear of blood clotting from the AstraZeneca vaccine, the failure of government to secure Pfizer vaccine when it was first available.

The bulk of opinions about vaccination has been arising from people residing in Sydney and Melbourne, the most populated two cities in Australia (Australian Bureau of Statistics, 2022 August [cited 24 August 2022](b)), with Melbourne slightly dominating in terms of negative stances and people from both cities displaying quite similar levels of positive emotions (Figure 3). Both cities have shown identical behavior in terms of the 5C scale, with negative attitudes attributed mostly to lack of confidence and calculative behaviour, while positive attitudes related mostly to constraints, collective responsibility and confidence behaviour. The prevalence of tweets and people's variable emotions in Melbourne and Sydney towards vaccination may also be related to long lockdown periods in Victoria and New South Wales where these two cities are located, respectively (Australian Bureau of Statistics, 2022 August [cited 24 August

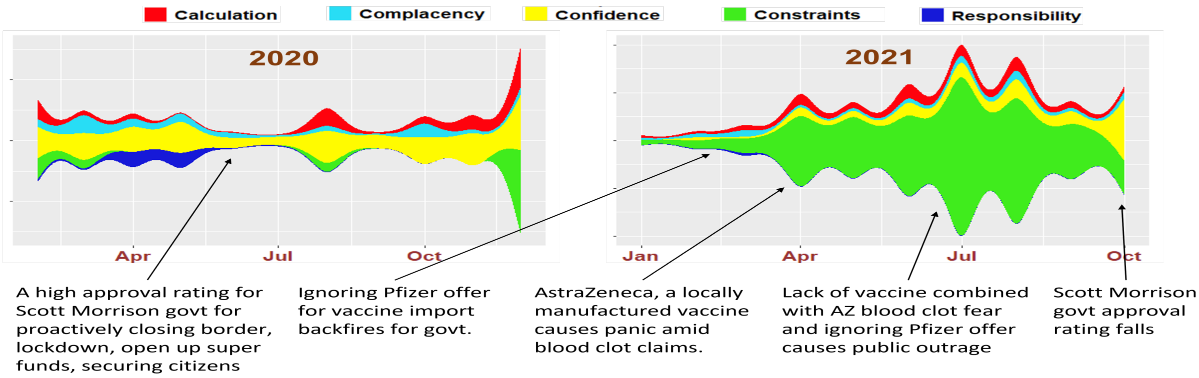

Figure 2: Changes of negative sentiment 5C's through 2020 and 2021 aligned with major COVID-related events. The description of the events are from: Statista (2020 August [cited 24 August 2022]); 9News (2021 August [cited 24 August 2022]); BBC News (2021 August [cited 24 August 2022](a),A); Reuters (2022 August [cited 24 August 2022]).

2022](a)).

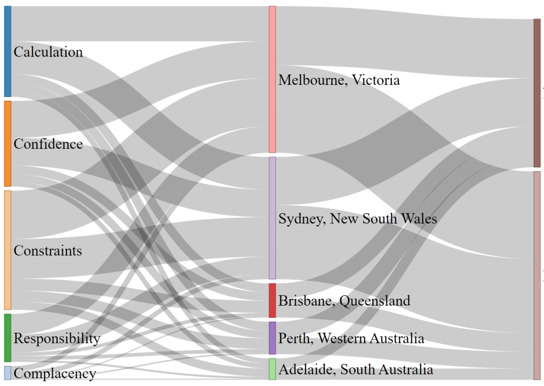

Figure 3: Distribution of positive and negative 5Cs based on location.

3.3 Vaccine behaviour identification

We perform three predictive modeling tasks in our analysis: (1) predicting the sentiment of tweets, (2) predicting the 5C vaccination behaviour in tweets, (3) predicting the combined sentiment and 5C categories in tweets:

Task 1 - Sentiment Identification: The first task is a binary classification problem to identify regardless of the type of behaviour if the post expresses positive or negative attitude towards vaccination. This task is framed as a binary classification problem.

Task 2 - 5C Categorisation: In task 2 our goal is to identify one of the 5C categories from the tweet content regardless of its sentiment. This task is useful to practitioners where there is no sentiment expressed, however, one of the 5C vaccine behaviours are expressed through posts. For example, a vaccine researcher might be expressing scientific data regarding the risks versus benefits of a vaccine in certain population groups without necessarily expressing an opinion. This task is formulated as a multi-class classification trained using categorical cross-entropy loss.

Task 3 - Combined Sentiment and 5C Categorisation : In this task, we categorise a post into one of the ten categories resulting from the two sentiments and the five vaccine behaviour types (10 classes). We formulate this task as a multi-class problem. To identify if the information in the sentiment and the vaccine attitudes are complementary, we compare the results with a model where Task 1 and 2 are trained separately but the results are combined after prediction.

For our benchmark models, we use BERT (Devlin et al., 2018) using the base-uncased English version. To make sure the results are reasonable, we compare Task 3 with a Gradient Boosting Machine (GBM) (Hastie et al., 2009) baseline using TF-IDF features. Prior to fitting the models, we pre-processed the tweet texts by removing stop words (only for GBM), tokenizing sentences and encoding the words to integers. We then randomly partitioned the dataset into train/test/validation sets in the ratio $7 : 1 : 2$.

GBM, in Task 3, is a tree-based ML model which sequentially fits new models to enhance the accuracy of the estimated response variable supervised learning tasks such as the classification problems we address here. For each predictive modeling task, we use 50 trees, each with a maximum depth of

Label	P%	R%	F1%
+Sentiment	90	87	**88**
-Sentiment	81	84	82
Confidence	48	55	51
Constraints	79	79	**79**
Complacency	25	18	21
Calculation	61	60	60
Collective Resp.	62	52	57
Combined at prediction	39	39	39

Table 6: Performance of BERT in predicting the sentiment (task 1), 5C categories (task 2) and combined sentiment and 5C during prediction.

A/B	C1	C2	C3	C4	C5
C1	297	84	39	166	70
C2	33	482	11	26	30
C3	2	5	3	1	1
C4	86	19	12	274	27
C5	44	24	7	14	138

Table 7: Confusion matrix for predicting the 5C categories (task 2).

Label	P%	R%	F1%
-Confidence	49	44	46
-Constraints	78	83	**80**
-Complacency	26	29	27
-Calculation	49	40	44
-Responsibility	20	13	16
+Confidence	47	51	49
+Constraint	33	28	30
+Complacency	7	10	8
+Calculation	61	57	**59**
+Collective Resp.	62	57	**59**

Table 8: Performance of BERT in predicting the combined sentiment and 5C categories (task 3).

15 and the minimum number of observations at each leaf node being also 15. The learning rate and column sample rate have been set to 0.1 and 0.4, respectively. We use grid search with 1 to 70 trees, having maximum tree depth in the range 3 to 7, with column sample rate of 0.4 to 1 and minimum rows from 1 to 100 to search for the best parameters in GBM using cross-validation. For BERT, we use the default parameters of the pre-trained model. Additionally, the number of nodes in the output layer equals the total number of classes to be predicted, the activation function of the output layer is "softmax", with the loss function being "categorical cross-entropy".

We evaluate the performance of the models using the precision, recall and F1 scores which are commonly applied measures in classification problems. Because of the imbalanced nature of our problem, we use macro-averaged F1 to evaluate across multiple classes.

4 Results

Our results assert that the largest F1 value corresponds to positive sentiments (88%) from the sentiment task (task 1) and the Constraint class (79%) in predicting the 5C task (task 2) (Table 6). Our results further show that BERT has correctly predicted the 5C categories 1191 times (Table 7).

For the combined sentiment and 5C prediction task (task 3), our prediction depicts that the negative Constraints (80%) and positive Calculation and Collective Responsibility categories (59% for both) have the highest F1 measure (Table 8). BERT has performed quite variably in predicting individual class in all three tasks due to the imbalanced nature of our dataset across the classes. This can be visualized from Figure 4, in which we can see that classes with lower proportions have been less likely predicted. As such, we fitted a GBM model to predict the combined sentiment and 5C task, and found that, overall, BERT performed better than GBM (F1 score of 48%).

We stress that GBM uses boosting, so it oversamples difficult instances, for example, less frequent ones, and thus can handle class imbalance in an indirect way to some extent. However, both GBM and BERT were to be affected by the class imbalance. When the weights of the low-frequency categories were increased to fix this issue, they reduced the accuracy of other categories, and hence we did not implement class weights in the models.

5 Conclusion

We presented the first large dataset for online monitoring vaccination behaviour in Australia, which is annotated by public sentiments towards vaccination and their psychological reasoning by the 5C scale. Our analysis has shown a close correlation between the sentiments of the tweets and the prominent events during the pandemic. Our analysis of vaccination behaviour from this dataset showed

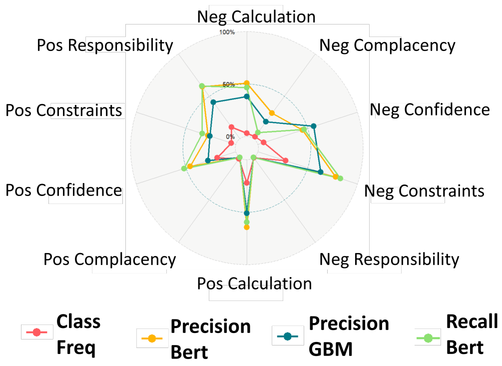

Figure 4: Comparison of precision and recall scores from the sentiment and 5C prediction task (task 3) with the distribution of 5C categories.

Model	P%	R%	F1%
BERT	53	38	**48**
GBM	36	28	32

Table 9: Comparison of the two benchmark methods for the combined sentiment and 5C tweet classification task.

large amount of negative emotions towards vaccination due to the constraints related to vaccine rollout caused by its shortage, delay in securing Pfizer vaccine when it was available, administering vaccination to aged care facilities, and the government's handling of the vaccination program.

Using this dataset, we predicted the sentiments towards vaccination, the 5C behaviour and the combined sentiment and 5C behaviour using BERT. All these predictive tasks are based on classification models, and showed variable performance across classes due to the imbalance proportion of data across classes. We compared performance of BERT with GBM in predicting the combined sentiment and the 5C categories and found that BERT has performed better than GBM.

Our work provides a proof of concept in the application of 5C scales to monitor vaccination behaviour using social media and can be extended to other domains such as Facebook and Google Trends.

6 Acknowledgements

This work is funded by The University of Queensland's AI4Pandemic group whose focus is the application of data science, machine learning, statistical learning, applied mathematics, and computation for managing pandemics both at the epidemic and clinical level. We are thankful for their support and hope that this work contributes to the aims of the group.

References

9News. 2021 August [cited 24 August 2022]. Labor accuses Morrison government of turning its back on Pfizer representatives in June 2020.

Büşra Akarsu, Dilara Canbay Özdemir, Duygu Ayhan Baser, Hilal Aksoy, İzzet Fidancı, and Mustafa Cankurtaran. 2021. While studies on COVID-19 vaccine is ongoing, the public's thoughts and attitudes to the future COVID-19 vaccine. *International journal of clinical practice*, 75(4):e13891.

Roy M Anderson, Carolin Vegvari, James Truscott, and Benjamin S Collyer. 2020. Challenges in creating herd immunity to sars-cov-2 infection by mass vaccination. *The Lancet*, 396(10263):1614–1616.

Australian Bureau of Statistics. 2022 August [cited 24 August 2022](a). Impact of lockdowns on household consumption - insights from alternative data sources.

Australian Bureau of Statistics. 2022 August [cited 24 August 2022](b). Regional population Statistics about the population for Australia's capital cities and regions.

Australian Government: Department of Health and Aged Care. 2022 August [cited 14 August 2022]. COVID-19 vaccination daily rollout update.

BBC News. 2021 August [cited 24 August 2022](a). Covid: Trigger of rare blood clots with AstraZeneca jab found by scientists.

BBC News. 2021 August [cited 24 August 2022](b). What's gone wrong with Australia's vaccine rollout?

Cornelia Betsch, Katrine Bach Habersaat, Sergei Deshevoi, Dorothee Heinemeier, Nikolay Briko, Natalia Kostenko, Janusz Kocik, Robert Böhm, Ingo Zettler, Charles Shey Wiysonge, et al. 2020. Sample study protocol for adapting and translating the 5c scale to assess the psychological antecedents of vaccination. *BMJ open*, 10(3):e034869.

Cornelia Betsch, Philipp Schmid, Dorothee Heinemeier, Lars Korn, Cindy Holtmann, and Robert Böhm. 2018. Beyond confidence: Development of a measure assessing the 5c psychological antecedents of vaccination. *PloS one*, 13(12):e0208601.

Jean-Christophe Boucher, Kirsten Cornelson, Jamie L Benham, Madison M Fullerton, Theresa Tang, Cora Constantinescu, Mehdi Mourali, Robert J Oxoby, Deborah A Marshall, Hadi Hemmati, et al. 2021. Analyzing social media to explore the attitudes and behaviors following the announcement of successful covid-19 vaccine trials: infodemiology study. *JMIR infodemiology*, 1(1):e28800.

Harshita Chopra, Aniket Vashishtha, Ridam Pal, Ananya Tyagi, Tavpritesh Sethi, et al. 2021. Mining trends of covid-19 vaccine beliefs on twitter with lexical embeddings. *arXiv preprint arXiv:2104.01131*.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Rachael H Dodd, Erin Cvejic, Carissa Bonner, Kristen Pickles, Kirsten J McCaffery, Julie Ayre, Carys Batcup, Tessa Copp, Samuel Cornell, Thomas Dakin, et al. 2021. Willingness to vaccinate against COVID-19 in Australia. *The Lancet Infectious Diseases*, 21(3):318–319.

Eve Dubé and Noni E MacDonald. 2020. How can a global pandemic affect vaccine hesitancy? *Expert review of vaccines*, 19(10):899–901.

Ben Edwards, Nicholas Biddle, Matthew Gray, and Kate Sollis. 2021. Covid-19 vaccine hesitancy and resistance: Correlates in a nationally representative longitudinal survey of the australian population. *PloS one*, 16(3):e0248892.

Eden Engel-Rebitzer, Daniel Camargo Stokes, Alison Buttenheim, Jonathan Purtle, and Zachary F Meisel. 2021. Changes in legislator vaccine-engagement on twitter before and after the arrival of the covid-19 pandemic. *Human vaccines & immunotherapeutics*, 17(9):2868–2872.

Kimberly A Fisher, Sarah J Bloomstone, Jeremy Walder, Sybil Crawford, Hassan Fouayzi, and Kathleen M Mazor. 2020. Attitudes toward a potential SARS-CoV-2 vaccine: a survey of US adults. *Annals of internal medicine*, 173(12):964–973.

Daniel Freeman, Bao S Loe, Andrew Chadwick, Cristian Vaccari, Felicity Waite, Laina Rosebrock, Lucy Jenner, Ariane Petit, Stephan Lewandowsky, Samantha Vanderslott, et al. 2020. Covid-19 vaccine hesitancy in the uk: the oxford coronavirus explanations, attitudes, and narratives survey (oceans) ii. *Psychological medicine*, pages 1–15.

H Fues Wahl, B Wikman Erlandson, C Sahlin, M Nyaku, and G Bencina. 2022. Analysis of vaccine messages on social media (twitter) in scandinavia. *Human vaccines & immunotherapeutics*, 18(1):2026711.

Allyson J Gallant, Louise A Brown Nicholls, Susan Rasmussen, Nicola Cogan, David Young, and Lynn Williams. 2021. Changes in attitudes to vaccination as a result of the covid-19 pandemic: A longitudinal study of older adults in the uk. *PloS one*, 16(12):e0261844.

Hao Gao, Qingting Zhao, Chuanlin Ning, Difan Guo, Jing Wu, and Lina Li. 2021. Does the covid-19 vaccine still work that "most of the confirmed cases had been vaccinated"? a content analysis of vaccine effectiveness discussion on sina weibo during the outbreak of covid-19 in nanjing. *International Journal of Environmental Research and Public Health*, 19(1):241.

Sefater Gbashi, Oluwafemi Ayodeji Adebo, Wesley Doorsamy, Patrick Berka Njobeh, et al. 2021. Systematic delineation of media polarity on covid-19 vaccines in africa: computational linguistic modeling study. *JMIR medical informatics*, 9(3):e22916.

Federico Germani and Nikola Biller-Andorno. 2021. The anti-vaccination infodemic on social media: A behavioral analysis. *PloS one*, 16(3):e0247642.

Talita Greyling and Stephanié Rossouw. 2022. Positive attitudes towards covid-19 vaccines: A cross-country analysis. *PloS one*, 17(3):e0264994.

Janessa Griffith, Husayn Marani, Helen Monkman, et al. 2021. Covid-19 vaccine hesitancy in canada: Content analysis of tweets using the theoretical domains framework. *Journal of medical Internet research*, 23(4):e26874.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Amir Hussain, Ahsen Tahir, Zain Hussain, Zakariya Sheikh, Mandar Gogate, Kia Dashtipour, Azhar Ali, Aziz Sheikh, et al. 2021. Artificial intelligence–enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study. *Journal of medical Internet research*, 23(4):e26627.

Hyeju Jang, Emily Rempel, David Roth, Giuseppe Carenini, Naveed Zafar Janjua, et al. 2021. Tracking covid-19 discourse on twitter in north america: Infodemiology study using topic modeling and aspect-based sentiment analysis. *Journal of medical Internet research*, 23(2):e25431.

Jessica Kaufman, Kathleen L Bagot, Jane Tuckerman, Ruby Biezen, Jane Oliver, Carol Jos, Darren Suryawijaya Ong, Jo-Anne Manski-Nankervis, Holly Seale, Lena Sanci, et al. 2022a. Qualitative exploration of intentions, concerns and information needs of vaccine-hesitant adults initially prioritised to receive COVID-19 vaccines in Australia. *Australian and New Zealand Journal of Public Health*, 46(1):16–24.

Jessica Kaufman, Jane Tuckerman, and Margie Danchin. 2022b. Overcoming covid-19 vaccine hesitancy: can australia reach the last 20 percent? *Expert Review of Vaccines*, 21(2):159–161.

Kin On Kwok, Kin-Kit Li, Wan In Wei, Arthur Tang, Samuel Yeung Shan Wong, and Shui Shan Lee. 2021a. Influenza vaccine uptake, covid-19 vaccination intention and vaccine hesitancy among nurses:

A survey. *International journal of nursing studies*, 114:103854.

Stephen Wai Hang Kwok, Sai Kumar Vadde, and Guanjin Wang. 2021b. Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: machine learning analysis. *Journal of medical Internet research*, 23(5):e26953.

Jeffrey V Lazarus, Scott C Ratzan, Adam Palayew, Lawrence O Gostin, Heidi J Larson, Kenneth Rabin, Spencer Kimball, and Ayman El-Mohandes. 2021. A global survey of potential acceptance of a COVID-19 vaccine. *Nature medicine*, 27(2):225–228.

Marie Fly Lindholt, Frederik Jørgensen, Alexander Bor, and Michael Bang Petersen. 2021. Public acceptance of covid-19 vaccines: cross-national evidence on levels and individual-level predictors using observational data. *BMJ open*, 11(6):e048172.

Joanne Chen Lyu, Eileen Le Han, and Garving K Luli. 2021. Covid-19 vaccine–related discussion on twitter: topic modeling and sentiment analysis. *Journal of medical Internet research*, 23(6):e24435.

Noni E MacDonald et al. 2015. Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34):4161–4164.

Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.

M Mills, C Rahal, D Brazel, J Yan, and S Gieysztor. 2020. COVID-19 vaccine deployment: behaviour, ethics, misinformation and policy strategies. *London: The Royal Society & The British Academy*.

Qian Niu, Junyu Liu, Masaya Kato, Tomoki Aoyama, and Momoko Nagai-Tanima. 2022a. Fear of infection and sufficient vaccine reservation information might drive rapid coronavirus disease 2019 vaccination in japan: Evidence from twitter analysis. *medRxiv*.

Qian Niu, Junyu Liu, Masaya Kato, Yuki Shinohara, Natsuki Matsumura, Tomoki Aoyama, Momoko Nagai-Tanima, et al. 2022b. Public opinion and sentiment before and at the beginning of covid-19 vaccinations in japan: Twitter analysis. *JMIR infodemiology*, 2(1):e32335.

Elise Paul, Andrew Steptoe, and Daisy Fancourt. 2021. Attitudes towards vaccines and intention to vaccinate against COVID-19: Implications for public health communications. *The Lancet Regional Health-Europe*, 1:100012.

Haley E Randolph and Luis B Barreiro. 2020. Herd immunity: understanding covid-19. *Immunity*, 52(5):737–741.

Reuters. 2022 August [cited 24 August 2022]. Australia PM's ratings tumble to lowest levels in nearly two years, poll shows.

Anthea Rhodes, Monsurul Hoq, Mary-Anne Measey, and Margie Danchin. 2021. Intention to vaccinate against covid-19 in australia. *The Lancet Infectious Diseases*, 21(5):e110.

Neeti Rustagi, Yachana Choudhary, Shahir Asfahan, Kunal Deokar, Abhishek Jaiswal, Prasanna Thirunavukkarasu, Nitesh Kumar, and Pankaja Raghav. 2022. Identifying psychological antecedents and predictors of vaccine hesitancy through machine learning: A cross sectional study among chronic disease patients of deprived urban neighbourhood, india. *Monaldi Archives for Chest Disease*.

Malik Sallam. 2021. COVID-19 vaccine hesitancy worldwide: a concise systematic review of vaccine acceptance rates. *Vaccines*, 9(2):160.

Holly Seale, Anita E Heywood, Julie Leask, Meru Sheel, David N Durrheim, Katarzyna Bolsewicz, and Rajneesh Kaur. 2021. Examining Australian public perceptions and behaviors towards a future COVID-19 vaccine. *BMC Infectious Diseases*, 21(1):1–9.

Susan M Sherman, Louise E Smith, Julius Sim, Richard Amlôt, Megan Cutts, Hannah Dasch, G James Rubin, and Nick Sevdalis. 2021. COVID-19 vaccination intention in the UK: results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. *Human vaccines & immunotherapeutics*, 17(6):1612–1621.

Statista. 2020 August [cited 24 August 2022]. COVID-19 and Leader Approval Ratings.

Linda Thunström, Madison Ashworth, David Finnoff, and Stephen C Newbold. 2021. Hesitancy toward a covid-19 vaccine. *Ecohealth*, 18(1):44–60.

Mallory Trent, Holly Seale, Abrar Ahmad Chughtai, Daniel Salmon, and C Raina MacIntyre. 2022. Trust in government, intention to vaccinate and covid-19 vaccine hesitancy: A comparative survey of five large cities in the united states, united kingdom, and australia. *Vaccine*, 40(17):2498–2505.

Junze Wang, Ying Zhou, Wei Zhang, Richard Evans, Chengyan Zhu, et al. 2020. Concerns expressed by chinese social media users during the covid-19 pandemic: content analysis of sina weibo microblogging data. *Journal of medical Internet research*, 22(11):e22152.

Jeremy K Ward, Caroline Alleaume, Patrick Peretti-Watel, Valérie Seror, Sébastien Cortaredona, Odile Launay, Jocelyn Raude, Pierre Verger, François Beck, Stéphane Legleye, et al. 2020. The French public's attitudes to a future COVID-19 vaccine: The politicization of a public health issue. *Social science & medicine*, 265:113414.

Annelot Wismans, Roy Thurik, Rui Baptista, Marcus Dejardin, Frank Janssen, and Ingmar Franken. 2021. Psychological characteristics and the mediating role of the 5c model in explaining students' covid-19 vaccination intention. *PloS one*, 16(8):e0255382.

World Health Organization. 2019 January [cited 14 August 2022]. Ten threats to global health in 2019.

Jia Xue, Junxiang Chen, Ran Hu, Chen Chen, Chengda Zheng, Yue Su, and Tingshao Zhu. 2020. Twitter discussions and emotions about the covid-19 pandemic: Machine learning approach.

Samira Yousefinaghani, Rozita Dara, Samira Mubareka, Andrew Papadopoulos, and Shayan Sharif. 2021. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, 108:256–262.