New Zealand COVID Tracer App: understanding usage and user sentiments

Marion Lara Tan
Massey University
M.L.Tan@massey.ac.nz

Oshada Senaweera
University of Colombo
oshada@stat.cmb.ac.lk

Asanka Gunawardana
University of Moratuwa
asankag@uom.lk

Mohamed Rasith
University of Moratuwa
mohamedrasith.18@business.mrt.ac.lk

Mohamed Suaib
University of Moratuwa
msuhaib.18@business.mrt.ac.lk

Theepika Shanthakumar
University of Moratuwa
theepika.18@business.mrt.ac.lk

Miznath Hisham
University of Moratuwa
miznath.18@business.mrt.ac.lk

ABSTRACT

The NZ COVID Tracer app is a part of Aotearoa New Zealand (NZ) Government's strategy to manage the COVID-19 pandemic. This paper investigates people's usage and sentiment on the app from its release in May 2020 to the end of 2021. Descriptive analysis of app data and sentiment analysis on user review data were used. The results show that before March 2021, the overall sentiment on the app was negative but gradually improved over time. The passive Bluetooth-tracing feature is utilised more consistently than the manual features. However, the increased proportion of positive sentiments is seen to increase with active app use. Results highlight the consistency of the Bluetooth-tracing feature but do not discredit the importance of manual interaction, as active use can improve the perception of the app. Insights from this study will be helpful as apps adapt to the changing context of the ongoing COVID-19 pandemic.

Keywords

Mobile apps, COVID-19 response, sentiment analysis, user reviews, app store data

INTRODUCTION

The NZ COVID Tracer app is a part of the Aotearoa New Zealand (NZ) Government's strategy to manage the COVID-19 pandemic in the country. The app's design and use have evolved through the different stages of the pandemic. This paper presents an investigation of the NZ COVID Tracer app performance from its release in May 2020 to the end of 2021.[1]

The NZ population of 5 Million has responded well to the app initiative, with approximately 3,500,000 total app registrations as of 31 Dec 2021 (Ministry of Health – Manatū Hauora, n.d.). Despite the high app registrations, the public's actual use of the app's functions may not show equivalence in numbers. The highest record on the number of active devices (where the user uses the device to perform manual features such as scanning or diary entries) in a day was only 1,450,129 as of 31 Dec 2021 (Ministry of Health – Manatū Hauora, n.d.). Usage patterns can also fluctuate depending on circumstances. This paper aims to uncover the performance of the NZ COVID Tracer app from its release in May 2020 to the end of 2021 and to provide an understanding of the public's perceptions of the app through user reviews. The paper seeks to answer the question: How did the public use and perceive the app during the period from May 2020 to the end of 2021?

[1] This paper covers this period of the pandemic before the Omicron variant spread in NZ.

BACKGROUND

NZ COVID Tracer app was launched in May 2020 by NZ's Ministry of Health to enable faster contact tracing for COVID-19 community cases. The app lets a person keep a digital diary of places visited by scanning official QR codes located on-premises or manually entering the details. The apps, starting 10 Dec 2020, had enabled a Bluetooth feature where the app conducts virtual handshaking of other nearby app-Bluetooth-enabled mobile devices. The Ministry of Health would then use the data from the virtual diary and Bluetooth handshakes for contact tracing should the user contract COVID-19 or would have been in contact with a person with COVID-19. The NZ COVID Tracer app developed its features through time, also in the changing context of the COVID-19 pandemic. Table 1 shows a summary of developments of the NZ COVID Tracer app.

Table 1. Summary of events relating to the NZ COVID Tracer app.

Date	Event
20 May 2020	Official launch of the NZ COVID Tracer app.
30 Jul 2020	The app was updated with an added feature that allowed users to add manual entries to their digital diary (on top of the QR code scanning feature).
19 Aug 2020	The Government made it compulsory for businesses to display the official app-compatible QR codes for premise entry doors and reception areas.
3 Sep 2020	The Government made it compulsory for all public transport providers, including buses, trains, ferries, ride-share vehicles, and operators, to provide QR codes for passengers.
23 Nov 2020	The app was updated to no longer require users to sign-up for an account or set up a password to use the app. All existing users no longer need to sign in periodically.
10 Dec 2020	A Bluetooth feature was added to the app.
22 Aug 2021	The Government announced that record-keeping, including scanning with the COVID Tracer app or manual signing, will now be mandatory for most events and businesses at all alert levels in response to the detection of the Delta variant
** These events are selected from the timeline and resources from the NZ Ministry of Health website.*	

Two studies have examined the public use and user experience of the NZ COVID Tracer app (Gasteiger et al., 2021; Tretiakov & Hunter, 2021). But these studies were conducted at the pandemic's earlier stages (between July and November 2020) and before the app had Bluetooth capabilities (implemented in December 2020). The studies by Gasteiger et al. (2021) and Tretiakov & Hunter (2021) provide good starting insights into using the app. The survey by Gasteiger et al. (2021), which ran between July to September 2020, found that only 31% of the respondents reported using the app frequently, while 24% used it sometimes, 21% had installed the app but not used it, and 24% had not installed the app. Barriers to installation and use include app-specific concerns such as technical, privacy, and security issues and behavioural and social barriers such as users' forgetfulness and lack of support from businesses on using QR codes (Gasteiger et al., 2021). App usage also relies on the environmental context (Tretiakov & Hunter, 2021); users notice a decline in app use during lower 'Alert Levels.'[2] Results from Tretiakov & Hunter's (2021) study show the need to utilise the concept of civic responsibility as it will likely appeal to users during high-threat conditions and will encourage the use of the app (Tretiakov & Hunter, 2021).

However, these studies have gaps. Tretiakov & Hunter's (2021) study only interviewed 34 users. The experience of the 34 may not represent the experiences of the entire NZ COVID Tracer app user base. Gasteiger et al.'s (2021) study had more participants (n = 343); however, the app aspect of the study was only part of a broader COVID-19 Health and Stress survey. The NZ COVID Tracer app was only appended in the broader survey with (1) a close-ended question on the frequency of use and (2) an open-ended question on their decision to use/not use the app. Gasteiger et al. (2021, p. 9) noted the limitation of their study: 'text-based responses may have limited the depth of data provided.'

METHOD

Given these gaps from previous literature, this study tries to investigate the NZ COVID Tracer app's performance and to gain deeper perspectives from the broader user base by analysing large data sets. It is important to mention that analysing large volumes of user reviews makes it possible to draw inferences on app usability and user experiences than gathered through structured surveys (Gasteiger et al., 2021; Tretiakov & Hunter, 2021).

[2] From March 2020 to December 2021, NZ operated on a four-tier alert level system (New Zealand Government, n.d.). Level 4 is the highest level where COVID-19 is not contained, and measures include strict lockdown. The lower the levels, the lower risk of transmission of COVID-19 hence more relaxed restrictions as levels go down.

Data sources

This study aims to enhance the current understanding of app use by investigating two larger data sources with broader coverages. The data sources for this study involve (1) the NZ Ministry of Health of daily statistics on usage and (2) the user reviews from the Google and iOS app stores on the NZ COVID Tracer app.

Ministry data

The NZ Ministry of Health provides statistical information about app usage. This is published on the Ministry of Health website[3], and a spreadsheet in CSV format is updated daily. The spreadsheet contains data from 19 May 2020 with the variables detailed in Table 2.

Table 2. Data variables from the Ministry of Health on the NZ COVID Tracer app.

Variable	Description	Data available from
App registrations	The number of registrations on the specified date range	19 May 2020
Poster scans	The total number of scans on QR codes made by the app users within 24 hours	19 May 2020
Daily entries	The total number of manual entries inputted by all app users within 24 hours	7 Jul 2020
Active device count	The number of devices that have either scanned QR codes or added a manual diary entry within 24 hours	24 Jun 2020
Bluetooth tracing numbers	The number of unique devices that have Bluetooth enabled and participating in Bluetooth contact tracing.	12 Dec 2020

User reviews from Google Play Store and Apple App Store

This study also analyses feedback gained through NZ COVID Tracer app user reviews. App stores, such as Google Play and Apple Store, provide means for their users to provide feedback through ratings and reviews (McIlroy et al., 2015). Analysing a large number of user reviews makes it possible to draw inferences on user experience that may not be replicated through solicited means (e.g., interviews and structured surveys) (Gebauer et al., 2008; Hedegaard & Simonsen, 2013). Users provide feedback in app stores without a predefined structure, and they can give reviews in an open-ended format where they can give as much praise or complaints about the app (Palomba et al., 2015; Tan et al., 2020a). Furthermore, these user reviews are not just summaries but are self-reports that contain insights into the user experiences (Hedegaard & Simonsen, 2013). Several studies have analysed app store data to get findings that have led to improvements in many aspects, including requirements engineering, planning, software design, and security and testing (Martin et al., 2016). For example, a study by Tan et al. (2020a) on user reviews from app stores has provided insights into users' perceptions of what makes disaster apps usable.

Data variables available from the app stores include (1) date of the review, (2) rating – score of 1 to 5, and (3) content of the user review. Raw data scraped for this study for the NZ COVID Tracer app included 3,552 reviews from Google Play and 554 reviews from Apple Store. A sample of the app user reviews is shown in Table 3.

Data analysis

The contribution of the paper is twofold. Firstly, a descriptive analysis of the NZ Ministry of Health data using numerical and graphical methods is carried out to understand the usage and the changes in activities of the COVID Tracer app. Secondly, sentiment analysis on the app user reviews from Google Play Store and Apple Store is performed to study the user perception of the app from its public release to the end of 2021.

Descriptive analysis of the Ministry of Health data

Numerical and graphical methods were used for the Ministry of Health data. The initial analysis looked first into Cumulative app registrations – the sum of the daily app registrations from19 May 2020 to 31 Dec 2021. This number is essential as it shows the total app registration representing the maximum possible number of mobile phone units that have the app installed. This number is the base for understanding the usage of the app, assuming that each registration represents a member of the population in NZ to be a user of the NZ COVID Tracer app.

[3] https://www.health.govt.nz/covid-19-novel-coronavirus/covid-19-data-and-statistics/covid-19-nz-covid-tracer-app-data

Note that the actual number of app users may be lower than the cumulative app registrations, as the data does not account for the possibility of double counting. For example, a single person may have multiple phones, or a person can acquire a new phone and install the app while retiring an old unit, or users can uninstall and reinstall. The limitation of the data is that it does not differentiate these situations but only accounts for every app registration

Table 3. Sample reviews and ratings.

Date	Rating	Content of the review
04 Dec 2021	1	*Will not work forget it and delete*
06 Dec 2021	2	*Super slow and when you try and do manual entry it can never find the location.*
15 Dec 2021	4	*Works most of the time. My Covid Record keeps falling off. Not so easy to use when out if you have to log in to the app each time.*
18 Dec 2021	3	*Scanning works well. Ability to show vaccination status needs to be easier - rather than opening a new window/going to Google pay why can't it just be a tab on the bottom of the tracer home screen?*
19 Dec 2021	2	*Unreliable app. Doesn't always work and often needs lot of positioning attemps to get it to scan code.*
19 Dec 2021	4	*It works fine, but can take ages to load, which can be VERY annoying*
23 Dec 2021	3	*A decent app. It has a long load time which is why I'm taking 2 stars off but still very good. It would encourage people to scan if it came to fit bit so you can just turn on the app and scan but sadly this collaboration hasn't happened yet*
25 Dec 2021	5	*This is so good so we stay safe*
27 Dec 2021	1	*Disgusting app disappears on screen no fix offered have to find it in my emails stupid system. And to think that is supposed to be something we have paid the government to produce for us, and can't even get a simple app to work properly.*
29 Dec 2021	5	*Helps keeps everyone safe*

The data variables were then subjected to descriptive numerical analysis where the average, maximum and minimum observations and standard deviation were calculated. Additional indicators were computed to provide further insight into app usage. Indicators include determining the average daily scans and manual entries per active user. Proportions on how many active and Bluetooth-enabled devices were compared to the maximum number of app registration were also computed. See Table 4 for more details on these indicators. Graphical methods were also used to plot the data variables over time to see changes in activity regarding the NZ COVID Tracer app. Substantial observable variations in the trend line were investigated and compared to significant events.

Table 4. Indicators to understand app usage.

Indicators	Formula	Description
Daily scans per active device	$= \frac{\sum_1^n scan/active\ device}{n_{days\ of\ observation}}$	An active app user averages this number of poster scans in a day
Daily manual entries per active device	$= \frac{\sum_1^n manual\ entry/active\ device}{n_{days\ of\ observation}}$	An active app user averages this number of manual entries in a day
Average active use ratio	$= \frac{\sum_1^n \frac{active\ device}{cummulative\ app\ registrations}}{n_{days\ of\ observation}}$	The average proportion of active apps (i.e., apps where users actively use QR scanning or diary entries)
Average Bluetooth ratio	$= \frac{\sum_1^n \frac{Bluetooth\ enabled\ apps}{cummulative\ app\ registrations}}{n_{days\ of\ observation}}$	The average proportion of apps that have Bluetooth enabled

Sentiment analysis on user reviews

Sentiment analysis is one of the major topics of Natural Language Processing which is useful for detecting people's opinions, sentiments, behaviours, emotions, appraisals, or attitudes towards products or services, issues, events, or topics (Liu & Zhang, 2012). This study identified the users' sentiments towards the NZ COVID Tracer app using the user review data from the app stores.

It should be noted that the sentiments toward the app will change over time due to various reasons. For example, a new update to the app can cause an increase in the positive perception of the app (sometimes it might cause a negative perception too). Therefore, it is informative to track the sentiments towards the app over time. The sentiment analysis focused on investigating users' perception of the app through time from its release to the end of 2021. Once the sentiments over time are identified, significant fluctuation can be further analysed to identify events or scenarios that may have caused those.

For the sentiment analysis, the app user review data from May 2020 to the end of 2021 were collected from Google Play Store and Apple Store. Each user review includes (1) the date of the review, (2) the content of the review in text format, and (3) the user rating on a scale of 1 to 5. User rating is a good indicator of the user perception of the app. However, it is limited to just a number and may not reflect the true sentiment of the user (Maks & Vossen, 2013). Therefore, the sentiment analysis was performed on the user review content in text format. When a user writes a review on the app, it reflects their sentiment towards it. However, manually reading all the reviews and identifying the sentiment is highly cumbersome and may lead to very subjective results. Two people might read the same review and come to two different conclusions about the sentiment. To avoid subjective interpretation, we used the four most frequently used machine learning (ML) classifiers and one deep learning (DL) classifier to classify each review's sentiment.

Machine learning and deep learning classifiers

The goal of sentiment analysis, through ML and DL approaches, is to deal with labelled data and help to create models using supervised learning algorithms, namely, decision tree (DT) (Zharmagambetov et al., 2015), random forest (RF) (Breiman, 2001), K-nearest neighbours (KNN) (Mitchell, 1997), multinomial logistic regression (MLR) (Hastie et al., 2009), and bidirectional long short-term memory (LSTM) (Ketkar & Santana, 2017). These are briefly explained below.

- DT is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome. The DT algorithm aims to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.
- RF utilizes ensemble learning, a technique that combines many classifiers to solve complex problems. An RF algorithm consists of many decision trees. The 'forest' generated by the RT algorithm is trained through bagging or bootstrap aggregating.
- KNN solves the classification problem by assigning the object to a class by a plurality vote from its *k* (a positive integer) neighbours. The algorithm captures the idea of similarity amongst the object with respect to its neighbours in terms of distance, proximity, or closeness.
- MLR is a linear algorithm, and the underlying technique is similar to linear regression. The term 'logistic' is taken from the logit function that is used in classification. The idea is to develop a model that best describes the relationship between the outcome and a set of independent variables.
- LSTM is an updated version of Recurrent Neural Network which handles log sequence dependencies well. The LSTM network is fed by input data from the current time instance and output of hidden layers from the previous time instance. These two data pass through various activation functions and valves in the network before reaching the output.

Labelled training data

A challenge in the sentiment analysis through ML and DL techniques is that they require labelled training data - data consisting of reviews and the corresponding sentiments, which are not available in the NZ COVID Tracer app user reviews data. Moreover, DL requires large amounts of data to learn parameters (Pan et al., 2020). Therefore, three data sets are labelled with sentiments: (1) Google Play Store apps user reviews from Kaggle, (2) Nykaa app reviews from Kaggle., and (3) Google Play Store and Apple Store user reviews of MySejahtera COVID tracer app in Malaysia have been collected to train the models. All data sets consist of user reviews of different apps in text format together with the sentiment labels as positive, negative, and neutral. Table 5 shows the distribution of all reviews after merging the three data sets. There are 32,189 experimental data including 17,135 labelled as positive sentiments, 10,007 as negative sentiments, and 5,047 labelled as neutral sentiments accounting for 53.23%, 31.09%, and 15.68%, respectively.

Table 5. The distribution of positive, negative, and neutral reviews of the merged data set.

	Positive reviews	Negative reviews	Neutral reviews
Count	17,135	10,007	5,047
Proportion	53.23%	31.09%	15.68%

Data pre-processing included removing punctuations, changing the review content into lower case, tokenisation, filtering stop words, and lemmatisation. After pre-processing, the review data were transferred into a feature vector representation using TF-IDF (Term Frequency-Inverse Document Frequency) vectoriser (Tripathy et al., 2016). Next, the data set was divided into two sets, 80% of the data as a training set and 20% as a testing set. The training set was used to train the models, and cross-validation was used to optimise the hyperparameters of each model class. Once an optimal model from each model class was obtained, the testing data was used to identify the performance of each model. The last two steps of the sentiment analysis are (1) to use the optimal model to predict the sentiment of the user reviews of the NZ COVID Tracer app and (2) to validate the results. The user rating of each review on a scale of 1 to 5 was used in the model validation. The platform used was Python in the Google Colab cloud service, and the sentiment analysis code is available on GitHub[4]. The results of the sentiment analysis are given in the Results section below.

FINDINGS/RESULTS

Descriptive analysis results

From the app launch in 2020 until 31 Dec 2021, there were 3,531,037 cumulative app registrations – the maximum number of possible apps installed. The number indicates that there is high app uptake in NZ (approximately 75% of a population of 5 Million); if it is assumed that each app registration translates to one person having the app.

Table 6 provides a descriptive numerical summary of the Ministry of Health Data. On any given day, there are 1,086,339 scans and 37,648 manual entries from all the app users. However, on average, there are only 554,360 users that actively use their apps for either scanning or manual entries per day. On average, 1,559,428 devices in a day have enabled Bluetooth contact tracing. These results indicate that the passive feature of the app (Bluetooth contact tracing) is utilised more than the user-active features of the app (i.e., scanning and diary entries).

Table 6. Descriptive numerical summary of Ministry of Health data variables.

	Average	Max	Min	Standard Dev.
App registrations per day	6,184	233,200	302	17,850
Scans per day (from all users)	1,086,339	3,962,782	1,407	895,366
Manual entries per day (from all users)	37,648	255,338	5	31,562
User active devices per day	554,360	1,450,129	4,800	373,932
Bluetooth-enabled devices per day	1,559,428	2,401,547	351,430	562,272

On average, there are 6,184 registrations per day during the study period. But the registrations are not constant throughout this period, as seen by the standard deviation of 17,850 registrations, maximum observation of 233,200 registrations in a day and minimum observation of 302 registrations a day. Prominent peaks appear when plotting the app registrations over time (See Figure 1).

An uptick in registrations was observed during the app release, with 143,405 registrations on 20 May 2020. Registrations rapidly declined to a stable 2,000 registrations a day. Then a significant increase is observed in August. Similar patterns of rapid increase-decrease-stabilisation are seen throughout the study period. Upon investigation, sudden increases occur with significant events in NZ regarding COVID-related events. Summarised events below are gathered from the NZ Government's Unite Against COVID-19 website (n.d.-a) and RNZ (2021):

- **11 Aug 2020** - Four new cases of COVID-19 are recorded in the community; the first time for a community case since NZ moved to Alert Level 1,
- **24 Jan 2021** - NZ records the first COVID-19 community case since November 2020
- **14 Feb 2021** - Three new cases of COVID-19 are recorded in the community. Auckland moved to Alert Level 3 at 11:59 pm. The rest of NZ moved to Alert Level 2.
- **28 Feb 2021** - Auckland again goes into lockdown, moved back to Alert Level 3 due to further community cases. The rest of NZ moved to Alert Level 2.
- **17 Aug 2021** - All of NZ moved to Alert Level 4 at 11:59 pm, due to detection of Delta in the community.

[4] https://github.com/theeps/New-Zealand-Covid-Tracer-App-Usage-and-Sentiments

- **23 Jun 2021** - A traveller from Sydney to Wellington tested positive for COVID-19 upon their return to Australia. Wellington moves to Alert Level 2. The rest of NZ remained on Alert Level 1.

These results show that uptake for app registration relates to significant events. Similarly, high deviations are observed with the other data variables (see Table 6 for a summary). These results indicate fluctuations in the app's usage throughout the study period.

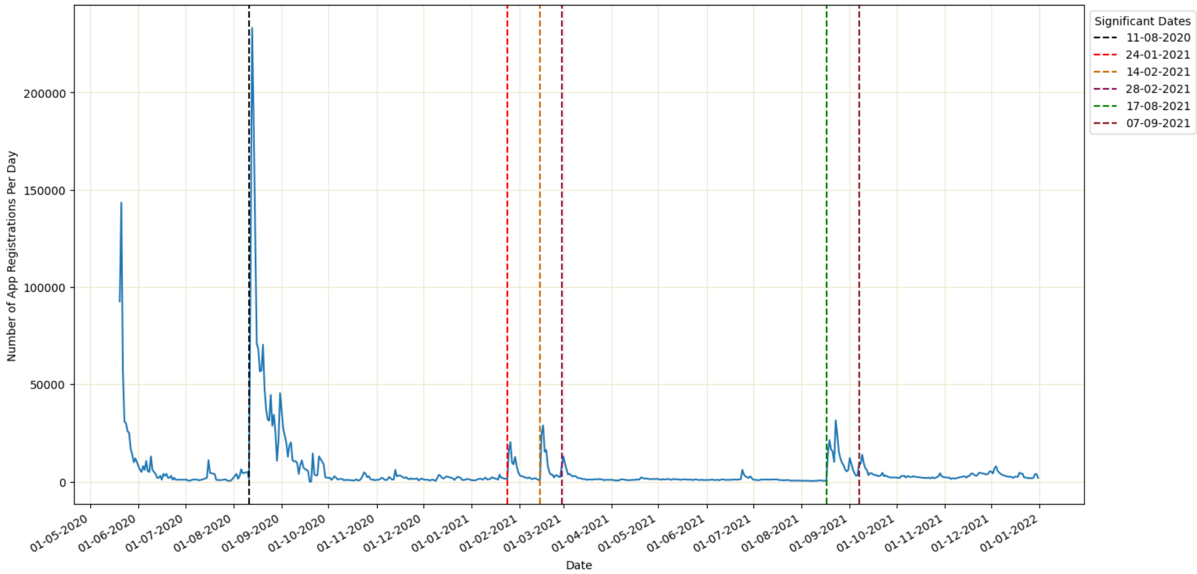

Figure 1. App registrations per day through the study period; vertical lines indicate dates associated with trend changes.

Figure 2 plots the number of scans recorded per day. It shows that scanning significantly increased after 11 Aug 2021, when the first community case was detected in NZ since it practically eliminated COVID-19 with zero community cases and went into Alert Level 1. Average scans before 11 Aug were approximately 26,000 scans a day. While average scans from 11 Aug to the end of 2020 were at 1 million scans a day, showing that people were scanning more often. Fluctuations in scans are seen throughout the period, but the scans in a day normalised to hundreds of thousands a day, not returning to pre-August 2020 levels. Another significant shift in scanning was observed on 7 Sep 2021, when NZ (except in the Auckland region) moved away from lockdowns down to Alert Level 2. As the population moved out of strict lockdown, more people were moving around and scanning. The average number of scans from 7 Sep to the end of 2021 was 2.6 Million scans per day. A significant drop can be seen on 25 Dec 2021, as most people stay home and most establishments close operations on Christmas day.

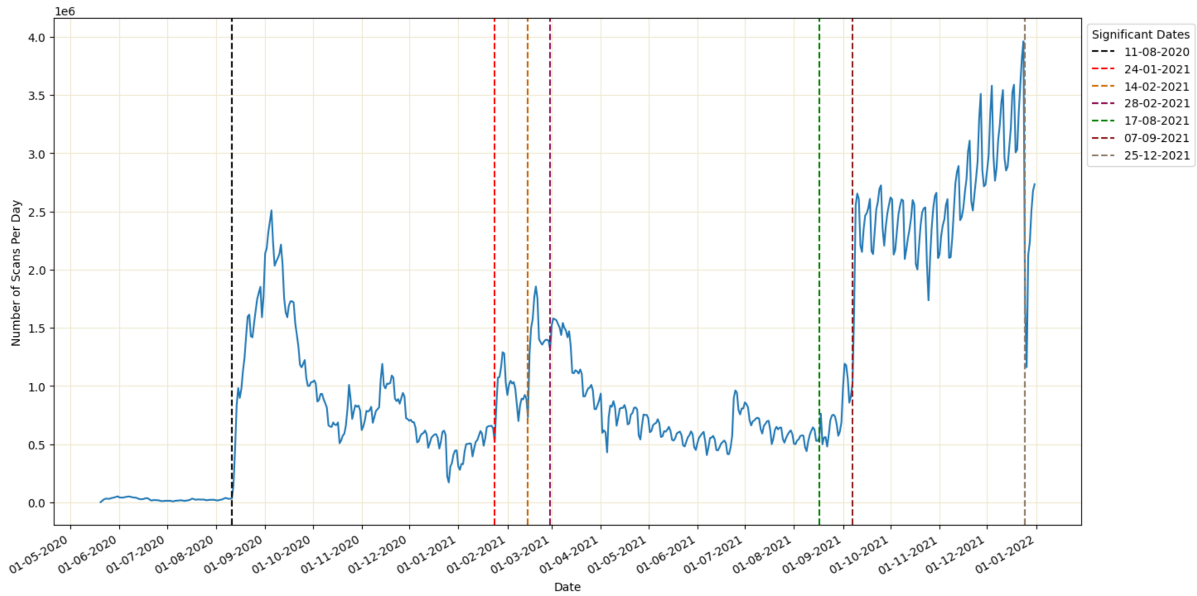

Figure 2. The number of scans per day through the study period; vertical lines indicate significant dates.

The computed indicators in Table 7 provide insight into app usage. An active device only scans 1.97 times a day. Users use the manual diary feature even less frequently, averaging only 0.09 entries per day per device. Of the possible total of approximately 3.5 million registered apps, only 20% actively engage with their devices. In comparison, 35% had their Bluetooth enabled during the study period.

Table 7. Indicators of app usage.

Indicator	Results
Daily scans per active device	1.97
Daily manual entries per active device	0.09
Average active use ratio	0.20
Average Bluetooth ratio	0.35

These utility ratios, however, may have changed over time. To investigate further, the cumulative app registrations, the number of user active devices, and the number of Bluetooth-enabled devices are plotted through time in Figure 3. The blue line represents the possible maximum in terms of the use of the app. The figure shows gaps between the maximum potential (blue line), the Bluetooth-enabled devices (green line), and the active devices (orange line). Although there are around 3.5 million devices with the NZ COVID Tracer app by the end of 2021, many devices are still not being used for scanning or diary entries. Also, a large number of devices did not have Bluetooth enabled.

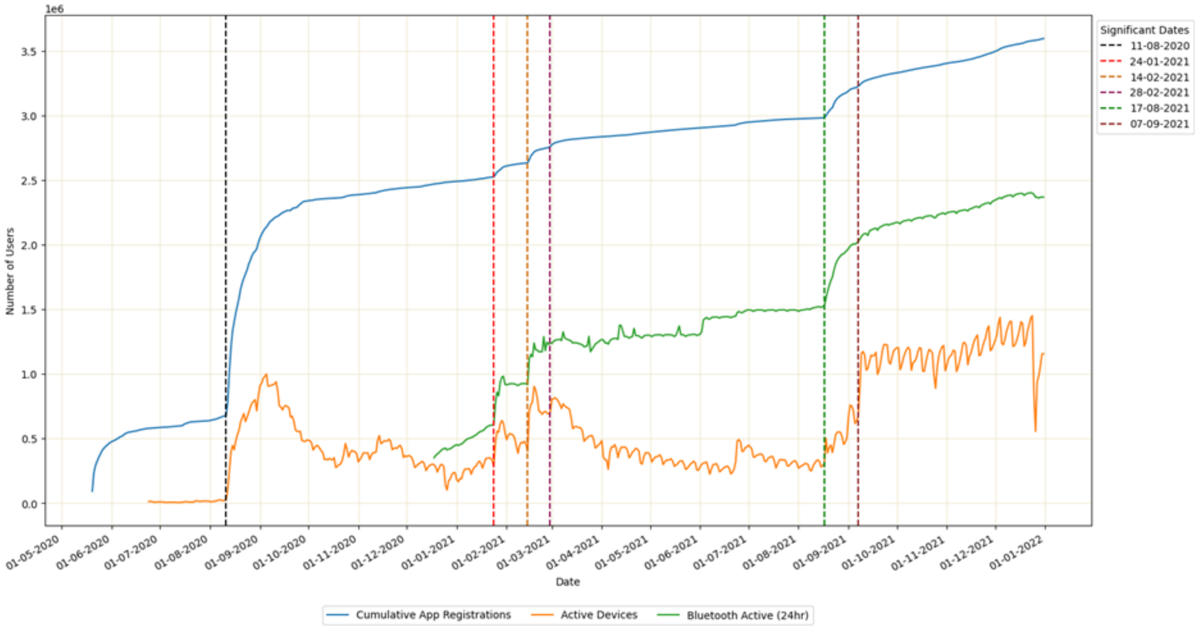

Figure 3. App utility through time. The blue line indicates the cumulative app registration through time. The orange line shows the active devices (i.e., scan or diary entry) per day. The green line portrays the Bluetooth-enabled devices per day through the study period; vertical lines indicate significant dates.

The 'active use ratio' (the active devices to cumulative registrations) and Bluetooth ratio (the number of Bluetooth-enabled devices over cumulative registrations) are plotted through time in Figure 4. The maximum value of the active ratio was attained at 0.474 on 4 Sep 2020 – the day after the Government made it compulsory for all public transport providers, including buses, trains, ferries, ride-share vehicles, and operators, to provide QR codes for passengers. After that peak, the active use ratio fluctuated but generally increased over time, especially over the later stages of 2021. The Bluetooth ratio did not fluctuate much in comparison and increased consistently over time. The Bluetooth feature seems to have a more consistent increase in use compared to the manual features of the app.

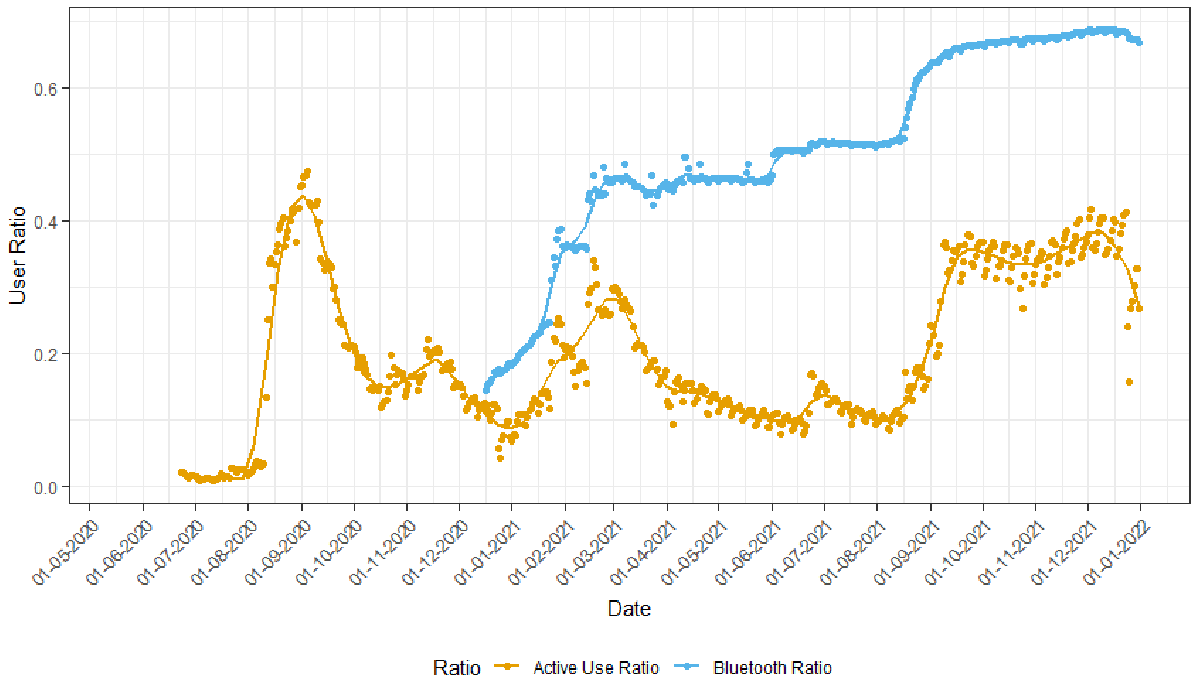

Figure 4. Active use ratio and Bluetooth ratio through time.

Sentiment analysis results

The study compares five classification models: RF, DT, KNN, MLR, and LSTM. Table 8 shows the classification results of each model, and Table 9 shows each classifier's accuracy. All these results represent how well each model performs on the testing data set.

Table 8. Classification results of RF, DT, KNN, MLR, and LSTM on performance metrics Precision (P), Recall (R), and F-measure.

	Positive			Negative			Neutral		
Model	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
RF	0.94	0.94	0.94	0.91	0.95	0.93	0.91	0.81	0.85
DT	0.93	0.93	0.93	0.91	0.92	0.91	0.83	0.81	0.82
KNN	0.82	0.90	0.86	0.86	0.60	0.70	0.62	0.78	0.69
MLR	0.92	0.95	0.94	0.92	0.94	0.93	0.91	0.77	0.83
LSTM	0.94	0.96	0.95	0.94	0.94	0.94	0.89	0.84	0.86

Table 9. Accuracy values obtained by each classifier.

Model	Results
RF	0.93
DT	0.91
KNN	0.79
MLR	0.92
LSTM	0.93

It can be readily seen that both RF and LSTM models performed equally well on the merged data set that consists of sentiment labels: positive, negative, and neutral in terms of accuracy (93%). From the performance metrics: Precision, Recall, and F-measure (P-R-F index), it can be seen that LSTM is more accurate in positive classification (P = 0.94, R = 0.96, F = 0.95) compared to RF (P = R = F = 0.94). Moreover, LSTM is more accurate in negative and neutral classification than RF in terms of F-measure. However, both models are almost equally accurate. Among the five classifiers, KNN has the least accuracy (79%) and the lowest P-R-F index in all three classifications. Hence, KNN does not perform well on the data set compared to the other four models. From the comparison of DT and MLR, MLR performed fairly well with an accuracy of 92% and F-measures of 0.94, 0.93, and 0.83 in positive, negative, and neutral classification, respectively. Overall, it can be concluded that both RF and LSTM models performed the best in accuracy and performance metrics. Thus, both models were used to predict the sentiments of the user reviews of the NZ COVID Tracer app.

Even though user ratings and the review's sentiment do not necessarily match (Maks & Vossen, 2013), they can be used to verify the results of the best models (RF and LSTM). Generally, a review rating of 4 or 5 indicates positive sentiment, a rating of 1 or 2 indicates negative sentiment, and a rating of 3 indicates neutral sentiment. Using this criterion, the sentiments predicted by both RF and LSTM were compared with those assigned by the user rating. This comparison validated the two models and computed the similarity between the two sentiments, where RF gave a similarity of 69.47%, and LSTM gave a similarity of 68.84%. The RF model was selected for further analysis of the sentiments, as it gives the highest similarity between the two sentiments.

Graphical methods were used to observe patterns and trends of the sentiments. Figures 5 and 6 show the changes in the proportions of three sentiments over time. These plots help to identify significant fluctuations of the sentiments (Figure 5) and can provide insight into how the perception of the app has changed during the study period (Figure 6).

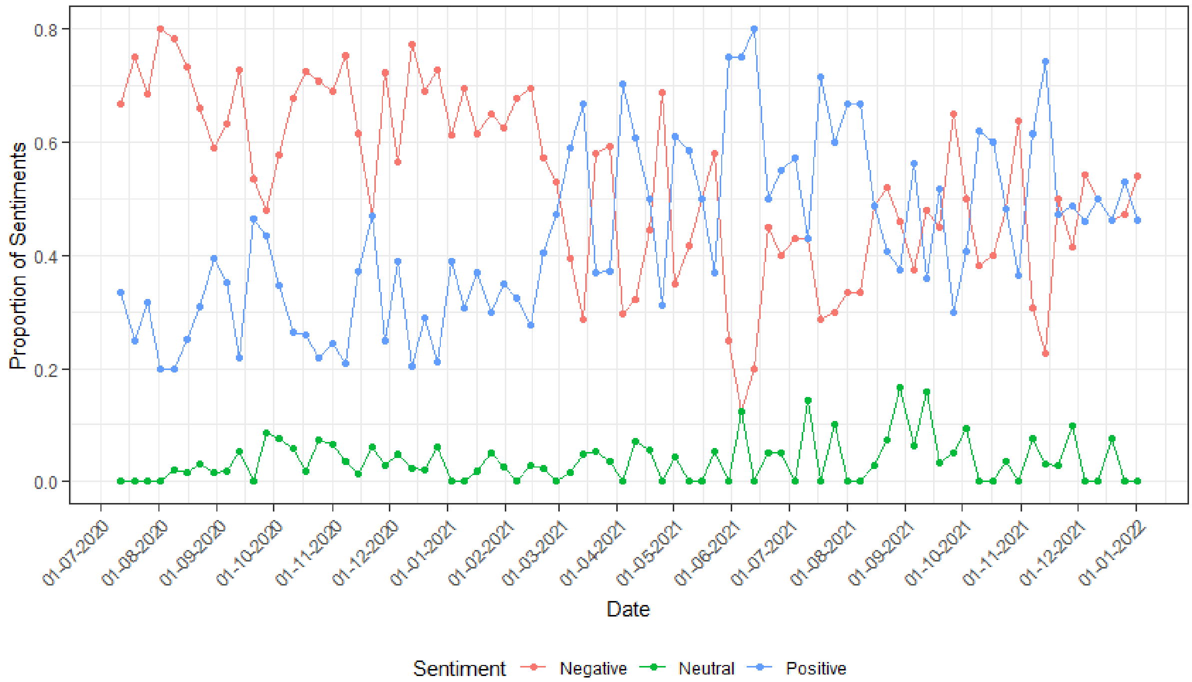

Figure 5. Weekly proportion of each sentiment.

Figure 5 plots the proportion of sentiments over a week through the study period. Proportion is calculated as the number of reviews for a sentiment category for a week divided by that week's total number of reviews. Before the first week of March 2021, the proportion of negative sentiments consistently outnumbered the proportion of positive sentiments. The week of 1 Mar 2021 shows a cross point between positive and negative sentiments; from that point, positive sentiments sometimes outweigh the negative sentiments. Fewer reviews on the week may have caused the fluctuation. However, upon investigating the history of COVID-19 events in NZ, the month of February 2021 saw numerous alert level changes in the country. On 28 Feb 2021, Auckland went into lockdown and moved back to Alert Level 3, while the rest of NZ moved to Alert Level 2. The change in sentiment may also coincide with the changing COVID-19 situation in NZ. This date was also noted in the increased app registrations, scanning, and Bluetooth use (see Figures 1, 2, and 3). In this case, the swing toward an increased proportion of positive sentiment coincides with the increased utility of the app. Future studies should investigate this relationship further to validate.

Another notable observation in Figure 5 is the increased gap between positive and negative sentiments from 25 May 2021 to June 2021. It should be noted that the Ministry of Health released the updated Version 5.0.0 of the app in May 2021. Communications on the Version 5.0.0 revealed that the update was made to make the app more useful and user-friendly; improvements included updates on the dashboard, digital diary, and Bluetooth tracing

(Ministry of Health - Manatū Hauora, 2021). The higher proportion of positive sentiments during this period may reflect how users generally supported the May 2021 app update.

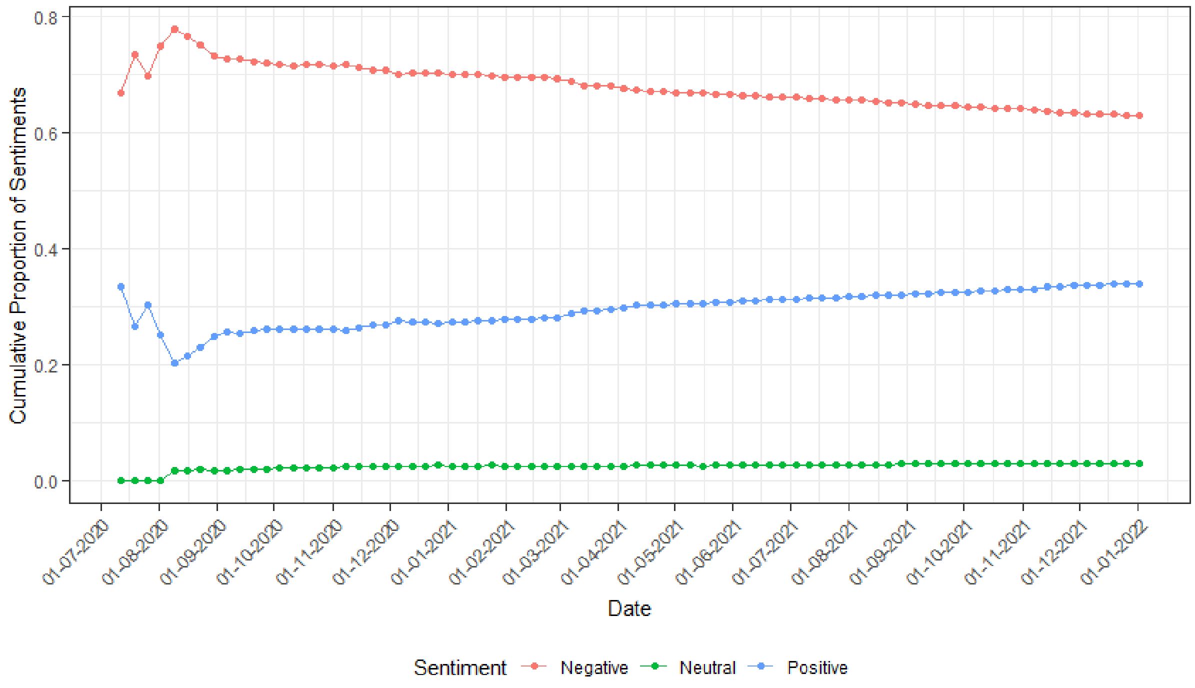

Figure 6. Cumulative proportion of each sentiment over time.

Figure 6 plots the cumulative proportion of each sentiment over time; this was conducted to investigate trends in the sentiments. Cumulative proportion is calculated as the total number of reviews for a sentiment category up to a specific date divided by the total number of reviews up to that date. Figure 6 indicates an upward trend of positive sentiments and a downward trend of negative sentiments, whereas neutral sentiments do not indicate any downward or upward trend. This implies that although the general sentiment of the user reviews is overall negative, user perception is improving over time.

DISCUSSION

The results from analysing the Ministry of Health data support the findings from Tretiakov & Hunter (2021) that app usage is affected by environmental context and the findings from Gastagier et al. (2021) that users only sometimes use the app. This study, however, provides further details on the NZ COVID Tracer app usage.

The result confirms that environmental context affects usage. However, lower alert levels do not necessarily result in a decline in usage. This study has shown that after the alert level was downgraded on 7 Sep 2021, the number of scans increased to an average of more than 2 million daily. This may be because lower alert levels mean more movement allowed for the population (from Alert Level 3 lockdown to return to business in Alert Level 2) hence an increase in the number of interactions between people and QR posters in establishments. Also, alert levels were lowered with active COVID-19 transmission still in the community. Hence there could have been an assumption of higher threat conditions which can encourage app use (Tretiakov & Hunter, 2021). This observation supports the need to utilise the concept of civic responsibility to encourage the app usage (Tretiakov & Hunter, 2021).

Manual use of the app is highly dependent on the environmental context. The active use ratio is not as high as the Bluetooth ratio, but this does not mean the manual features should be abandoned. There is good momentum in the usage of manual features as the active ratio trend is increasing through time despite fluctuations. Users should still be encouraged to interact with the app to establish trust despite limited interaction (Tan et al., 2020b). Even if apps are not regularly used, the users need to be able to interact and feel that the app is active and up-to-date or else it may result in users uninstalling the app completely (Tan et al., 2020b).

Results from this study highlight the importance of the Bluetooth feature of the app. The Bluetooth function is a passive feature that is, by default, enabled and will work in the background. Users can, however, choose to disable the Bluetooth feature. Despite being introduced later than the manual features of the app, there is a higher proportion of the registered apps that has Bluetooth enabled than active devices (i.e., users using the scan or diary feature). By the end of 2021, more than 2 million units have Bluetooth enabled, and their devices conduct virtual handshaking without the users' conscious knowledge. For contact tracing, it seems Bluetooth tracing should be utilised more, and more users should be encouraged to enable the Bluetooth feature. Bluetooth data and its utilisation for contact tracing should be investigated further.

Sentiment analysis of user reviews shows that many user reviews are either negative or positive sentiments rather than neutral sentiments. This is consistent with findings from other app store analyses, where satisfaction extremes are well represented (Hedegaard & Simonsen, 2013; Tan et al., 2020a). Because of the nature of user reviews, users tend to leave comments when they are highly satisfied or dissatisfied. The overall sentiment of the NZ COVID Tracer app is negative. However, the analysis conducted in this study looked at sentiment not just as an overarching whole but rather as the sentiments were plotted through time to observe any significant changes or trends. Extreme comments (e.g., positive or negative) may tend to dominate user reviews; hence, these are possibly ignored in understanding app performance. However, despite having a more significant proportion of negative sentiments for the NZ COVID Tracer app, plotting the sentiments' movements provided insight into specific periods of interest. It can be seen that negative sentiments are decreasing while positive sentiment is increasing, indicating sentiment is improving over time.

The findings suggest ways to help improve contact tracing by enhancing utilisation. Firstly, Bluetooth utilisation shows better consistency throughout the study period and does not depend on external events to prompt usage. Automated approaches such as Bluetooth should be prioritised, and efforts should be made to communicate and encourage people to turn the Bluetooth function on. Secondly, interaction with the app is also crucial, as observed through the sentiments over time. Improving people's positive perception of the app requires users also to be active (e.g. using QR code scanning and manual tracing). Although manual use is not as effective for contact tracing as the Bluetooth feature, as it is used less consistently, it still has its purpose as it gives users some agency over the app and may contribute to positive perceptions.

A limitation of this study is that only visual observations were analysed. Future studies can use inferential statistical methods to test for the significance of various data variables to gain further insight. Other ML and DL methods on textual data can also be used to investigate reasons for negative and positive sentiments. Future studies on machine learning can help further investigate user reviews. An area to explore is to find ways to extract usability insights for app improvements. Another limitation is that the study of this app only applies to the COVID-19 pandemic context. Further research is needed to investigate whether the utility of contact tracing apps can be used for other cases (e.g. flu or measles).

CONCLUSION

At the beginning of the paper, a research question was raised: how did the public use and perceive the app during the period from May 2020 to the end of 2021? The NZ public has high app uptake, but app utility can still improve. The number of active users (i.e., scanning and diary entry) fluctuates and is inconsistent, which may not be ideal for contact tracing. In contrast, there is a higher proportion of Bluetooth-enabled devices, and Bluetooth utility is consistently increasing over time. The consistency of the Bluetooth feature is promising as it may be more beneficial for contact tracing purposes. The sentiment analysis of the user reviews shows a higher proportion of negative sentiments on the app but, through time, showed improvement. Fluctuations favouring increased proportions of positive sentiments coincide with increased app utility in Feb-Mar 2021 and the app update in May 2021. This may imply that encouraging active use and improving usefulness will make users more satisfied with the app.

Insights from this study will be helpful as the changes in the COVID-19 pandemic continues. As of 2022, the NZ COVID Tracer app's use for the pandemic response has changed. With COVID-19 widespread in the community, the responsibility for contact tracing has moved to the people testing positive to inform their close contacts. The Government has since also dropped the mandatory scanning and displaying of QR code posters; however, the Bluetooth feature of the app is still being utilised, as alerts may still be sent out to your phone when pinged by the Bluetooth feature. Results from this study affirm the Bluetooth feature's utility but also highlight the need to understand user sentiments and how it coincides with people's active use of the app. Findings from this study show the importance of users still being encouraged to interact with the app. Even if the NZ COVID Tracer app is no longer used the same way as in 2020 and 2021, the users need to be encouraged to interact and have positive sentiments about the app. The danger is that with inactivity, users will uninstall the app entirely, which may be problematic when the NZ COVID Tracer app needs to be used for future contact tracing in response to the pandemic.

ACKNOWLEDGEMENTS

The authors acknowledge Massey the University Research Fund for supporting this research.

REFERENCES

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Gasteiger, N., Gasteiger, C., Vedhara, K., & Broadbent, E. (2021). The more the merrier! Barriers and facilitators to the general public's use of a COVID-19 contact tracing app in New Zealand. *Informatics for Health and Social Care*, 1–12. https://doi.org/10.1080/17538157.2021.1951274

Gebauer, J., Tang, Y., & Baimai, C. (2008). User requirements of mobile technology: results from a content analysis of user reviews. *Information Systems and E-Business Management*, *6*(4), 361–384. https://doi.org/10.1007/s10257-007-0074-9

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Hedegaard, S., & Simonsen, J. G. (2013). Extracting usability and user experience information from online user reviews. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 2089–2098. https://doi.org/10.1145/2470654.2481286

Ketkar, N., & Santana, E. (2017). *Deep learning with Python*. Apress.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 415–463). Springer. https://doi.org/https://doi.org/10.1007/978-1-4614-3223-4_13

Maks, I., & Vossen, P. (2013). Sentiment analysis of reviews: Should we analyze writer intentions or reader perceptions? *International Conference Recent Advances in Natural Language Processing, RANLP, September*, 415–419.

Martin, W., Jia, Y., Sarro, F., & Harman, M. (2016). *A survey of app store analysis for software engineering*.

McIlroy, S., Ali, N., Khalid, H., & Hassan, A. E. (2015). Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Software Engineering*, 1067–1106. https://doi.org/10.1007/s10664-015-9375-7

Ministry of Health - Manatū Hauora. (2021). *NZ COVID Tracer – May 2021 update (full length)*. YouTube. https://www.youtube.com/watch?v=5NA04Q1rYlI

Ministry of Health – Manatū Hauora. (n.d.). *COVID-19: NZ COVID Tracer app data*. Retrieved March 14, 2021, from https://www.health.govt.nz/covid-19-novel-coronavirus/covid-19-data-and-statistics/covid-19-nz-covid-tracer-app-data

Mitchell, T. M. (1997). *Machine Learning* (Vol. 1, Issue 9). McGraw-Hill.

New Zealand Government. (n.d.). *History of the COVID-19 Alert System*. Unite Against COVID-19. Retrieved 13 Jul 2022, from https://covid19.govt.nz/about-our-covid-19-response/history-of-the-covid-19-alert-system/

Palomba, F., Linares-Vasquez, M., Bavota, G., Oliveto, R., Di Penta, M., Poshyvanyk, D., & De Lucia, A. (2015). User reviews matter! Tracking crowdsourced reviews to support evolution of successful apps. *IEEE 31st International Conference on Software Maintenance and Evolution*, 291–300. https://doi.org/10.1109/ICSM.2015.7332475

Pan, Y., Chen, Z., Suzuki, Y., Fukumoto, F., & Nishizaki, H. (2020). Sentiment analysis using semi-supervised learning with few labeled data. *Proceedings - 2020 International Conference on Cyberworlds, CW 2020*, 231–234. https://doi.org/10.1109/CW49994.2020.00044

RNZ News. (2021). *Timeline: The year of Covid-19 in New Zealand*. RNZ Website. https://www.rnz.co.nz/news/national/437359/timeline-the-year-of-covid-19-in-new-zealand

Tan, M. L., Prasanna, R., Stock, K., Doyle, E. E. H., Leonard, G., & Johnston, D. (2020a). Modified usability framework for disaster apps: A qualitative thematic analysis of user reviews. *International Journal of Disaster Risk Science*. https://doi.org/10.1007/s13753-020-00282-x

Tan, M. L., Prasanna, R., Stock, K., Doyle, E. E. H., Leonard, G., & Johnston, D. (2020b). Understanding end-users' perspectives: Towards developing usability guidelines for disaster apps. *Progress in Disaster Science*, *7*, 100118. https://doi.org/10.1016/j.pdisas.2020.100118

Tretiakov, A., & Hunter, I. (2021). User Experiences of the NZ COVID Tracer App in New Zealand: Thematic Analysis of Interviews. *JMIR MHealth and UHealth*, *9*(9), e26318. https://doi.org/10.2196/26318

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, *57*, 117–126. https://doi.org/10.1016/j.eswa.2016.03.028

Zharmagambetov, A. S., Pak, A. A., Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2015). Sentiment Analysis of a document using deep learning approach and decision trees. *12th International Conference on Electronics, Computer, and Computation (ICECCO). IEEE*, 1–2.