Taylor & Francis
Taylor & Francis Group

Behaviour & Information Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tbit20

Maternal mental health monitoring in an online community: a natural language processing approach

Zhen Zhu

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2024.2333927

Maternal mental health monitoring in an online community: a natural language processing approach

Zhen Zhu

Kent Business School, University of Kent, Canterbury, UK

ABSTRACT
Digital maternity support communities are increasingly popular. The communities are often based on discussion forums called 'birth clubs', to which users are assigned according to their estimated due months. Distinguishing between support-seeking and non-support-seeking posts submitted to these 'birth clubs' is a crucial first step for monitoring maternal mental health. This study utilised natural language processing (NLP) techniques on 52,558 posts collected from one of the largest online maternity communities in China, employing machine learning algorithms trained for post classification with a randomly selected and manually labelled subset of 3000 posts. The results validated the properties of information similarity and time sensitivity within the post data, and demonstrated the feasibility of employing simple algorithms and small training sets for effective maternal mental health monitoring.

ARTICLE HISTORY
Received 17 August 2023
Accepted 13 March 2024

KEYWORDS
digital health; maternal mental health; natural language processing; machine learning

1. Introduction

Maternal mental health issues have emerged as a significant concern worldwide, affecting women during pregnancy and the postpartum period (Alderdice, McNeill, and Lynn 2013; Figueiredo and Conde 2011; Hermann, Fitelson, and Bergink 2021; Lee et al. 2007; Ross, McLean, and Psych 2006). The challenges faced by expectant mothers require adequate support to ensure their well-being and that of their newborns (Rahman et al. 2013; Wachs, Black, and Engle 2009). In recent years, digital technologies have revolutionised the way we connect and communicate, opening up new possibilities for addressing maternal mental health at a large scale. Among these solutions, online maternity support communities have gained global popularity by facilitating peer-to-peer connections and information exchange (Chivers et al. 2020; Denton et al. 2020; Jiang and Zhu 2022a, 2022b; Wexler et al. 2020).

The journey of pregnancy can be a roller coaster of emotions for expectant mothers, encompassing joy, anxiety, stress, and uncertainty. However, some women experience more serious mental health challenges, including depression, anxiety, and post-traumatic stress disorder during pregnancy and after childbirth. These conditions can have detrimental effects on both maternal well-being and infant development (Figueiredo and Conde 2011; Lee et al. 2007; Rahman et al. 2013; Ross, McLean, and Psych 2006; Wachs, Black, and Engle 2009).

Historically, expectant mothers relied on traditional support systems, such as family, friends, and healthcare providers, for advice and emotional support (Bedaso et al. 2021; Orr 2004; Orr and Miller 1997). While these systems remain valuable, they may not always be sufficient, especially for women who lack close family ties or live in isolated communities. Additionally, seeking support from people within the same local community may limit the diversity of experiences and insights available (Broadbent and Papadopoulos 2013; Drentea and Moren-Cross 2005).

Digital maternity support communities offer a novel approach to connecting expectant mothers, transcending geographical boundaries and temporal constraints. The anonymity, accessibility, and convenience offered by digital platforms have paved the way for individuals facing similar challenges to connect and share their experiences and concerns in a relatively safe and non-judgmental space (Gleeson, Craswell, and Jones 2019; Gui et al. 2017). These online maternity communities typically provide cohort-based discussion forums. For example, the discussion forms can be created on a monthly basis and be named as 'birth clubs', to which users are assigned according to their estimated due months (Jiang and Zhu

CONTACT Zhen Zhu z.zhu@kent.ac.uk Kent Business School, University of Kent, Sibson, Parkwood Road, Canterbury, Kent CT2 7FS, UK

2022a, 2022b). These dedicated discussion forums foster a sense of belonging and validation, reassuring users that they are not alone in their experiences.

User-generated content on these discussion forums presents an opportunity for leveraging text data to monitor mental health and bolster healthcare management practices. Through sophisticated natural language processing (NLP) techniques, these platforms offer insights into users' emotional states and mental well-being, aiding in the early detection of conditions like depression and anxiety. By identifying linguistic markers and patterns, healthcare professionals can intervene proactively, while also respecting ethical considerations regarding privacy and consent, ultimately fostering a data-driven, supportive approach to mental healthcare on a much broader scale (Drydakis 2021; Hinduja et al. 2022; Kim et al. 2020). Whereas there is an increasing body of research focussed on general mental health monitoring in social media platforms such as Twitter through text analysis (Coppersmith, Dredze, and Harman 2014; Di Cara et al. 2023; Hinduja et al. 2022; McClellan et al. 2017; Tommasel et al. 2022), there remains a notable gap in the literature concerning the specific domain of maternal mental health monitoring within online maternity communities.

Therefore, this study makes a distinctive contribution to the existing body of literature by introducing several novel aspects that broaden our understanding of monitoring maternal mental health within a digital platform.

First, this study takes a pioneering step by delving into the realm of maternal mental health monitoring within a real-world digital platform. The data was collected from one of the largest online maternity communities in China and in the world. Second, this study contributes to the literature by validating two pivotal properties of post data within the context of this online community, namely information similarity and time sensitivity. This validation provides valuable insights into the dynamics of information exchange and temporal factors that influence maternal mental health discussions. Third, this study justifies the feasibility of monitoring maternal mental health with relatively straightforward machine learning algorithms and small training sets. The results not only underscore the feasibility of utilising NLP techniques for mental health assessment but also emphasise the potential for scalability and cost-effective practical implementation.

The rest of the paper is organised as follows. Section 2 provides a literature review. Section 3 introduces the data, proposes the hypotheses, and suggests the classification methods based on NLP. Section 4 summarises the results before Section 5 concludes the paper.

2. Literature review

2.1. Online maternity communities

Online maternity communities have gained significant attention in recent years as platforms where expectant and new mothers gather to share information, seek support, and connect with others in similar life stages. These virtual spaces have become popular due to their accessibility, convenience, and potential for emotional support (Gleeson, Craswell, and Jones 2019; Jiang and Zhu 2022a, 2022b). Researchers have investigated various aspects of online maternity communities, including their impact on maternal well-being (Jiang and Zhu 2022b), and the dynamics of peer support (Jiang and Zhu 2022a).

For instance, previous studies have highlighted the importance of peer support in online maternity communities. Expectant and new mothers often turn to these platforms to share personal experiences, seek advice, and receive reassurance from others who are facing similar challenges. Online communities provide a sense of camaraderie and understanding that can mitigate feelings of isolation and anxiety (Gleeson, Craswell, and Jones 2019). Moreover, research has indicated a positive association between participation in online maternity communities and improved emotional well-being, especially during the COVID-19 pandemic (Silva-Jose et al. 2022; Zhou et al. 2021). Engaging in these communities can enhance a sense of empowerment, reduce stress, and promote self-efficacy among expectant and new mothers. However, it is important to note that excessive reliance on online support can also contribute to feelings of comparison and inadequacy (Adams 2008).

2.2. Online mental health monitoring

The availability of online data also enables the possibility of mental health monitoring. This approach enables early detection of mental health changes by continuously tracking user-generated data (Drydakis 2021; Hinduja et al. 2022; Kim et al. 2020). Previous studies have demonstrated the potential for identifying shifts in mood, sleep patterns, and behaviour through data collected from wearable devices, smartphone apps, and online questionnaires (Naslund, Aschbrenner, and Bartels 2016; Onyeaka et al. 2021).

Many online monitoring platforms leverage algorithms to provide personalised interventions based on users' data. These interventions may include tailored educational resources, coping strategies, and referrals to mental health professionals (D'Alfonso 2020; Van

Der Krieke et al. 2014). Meanwhile, research has highlighted the importance of user-centred design, gamification elements, and continuous feedback loops to maintain user interest and motivation (Brown et al. 2016; Sardi, Idri, and Fernández-Alemán 2017). Furthermore, privacy and security concerns surrounding the collection and storage of sensitive mental health data have been explored extensively. Researchers have emphasised the need for transparent data usage policies, secure data storage, and compliance with ethical guidelines (Bennett, Bennett, and Griffiths 2010; Parker et al. 2019).

While a growing body of research is dedicated to general mental health monitoring on social media platforms like Twitter, typically using text analysis (Coppersmith, Dredze, and Harman 2014; Di Cara et al. 2023; Hinduja et al. 2022; McClellan et al. 2017; Tommasel et al. 2022), a significant research gap persists when it comes to bridging the above two research strands, and to monitoring maternal mental health within online maternity communities in particular.

3. Data, hypotheses and methods

3.1. Background and data

The data used in this study was collected from one of the largest online maternity communities in China and also in the world. Its core service is to facilitate information exchange and social support about maternal caring between pregnant women (Jiang and Zhu 2022a, 2022b). As its key feature, the online community assigns pregnant users into peer groups (i.e. the so-called 'birth clubs') based on their estimated due months. For example, the users who are expecting to give birth on any day in March 2024 will be asked about their estimated due dates upon their registration to the online platform, and will be assigned to the 'Birth Club March 2024' accordingly. Therefore, new 'birth clubs' are created on a monthly basis with clearly defined target users, and these 'clubs' are basically online discussion forums where users can submit posts, which then can be responded to with comments. Like other online discussion forums, each 'birth club' has a few moderators who help manage and organise the online discussion when needed (e.g. promoting instructive or informative content or responding to issues reported by users).

To monitor the mental health of these maternal users, a natural starting point is to detect related signals from their posts. For a feasibility study, 52,558 posts were collected from three 'birth clubs' corresponding to the due months of March, April, and May 2018. Each 'birth club' was observed and its posts were

Table 1. Summary statistics.

Data	# Posts	Average # Characters
Birth Club March 2018	20,085	12.71
Birth Club April 2018	17,126	12.83
Birth Club May 2018	15,347	13.30
1st Trimester	18,168	12.25
2nd Trimester	14,599	12.34
3rd Trimester	19,791	13.97
All	52,558	12.92

collected for 10 months since its inception (i.e. a typical pregnancy duration). Note that there are two important dimensions of this dataset. One is that it covers three 'birth clubs' corresponding to the due months of March, April, and May 2018, and the other is that the posts were collected with time stamps and therefore each post can be arranged according to the pregnancy timeline (e.g. in one of the three trimesters). Table 1 summarises the number of posts and the average number of characters of posts for each of these two dimensions as well as for the whole dataset. May 2018 and the last trimester had slightly longer posts (13.30 and 13.97 characters respectively) than others on average. The number of posts was similar across each dimension. Among the 'birth clubs', the March 2018 had the most posts (38.2%). Among the trimesters, the third trimester had the most posts (37.7%).

Figure 1 shows the top 20 most frequent words in the posts. Note that each original Chinese word is accompanied by its English translation to the left. As one might expect, the most frequent words include 'baby', 'pregnant', 'belly', and more technical terms such as 'four-dimensional (ultrasound)' and 'foetal heart', which are clearly related to maternity. However, the posts also frequently involve commercially related words such as 'purchase' and '(product) review'. In other words, a post submitted to a 'birth club' can serve not only to ask questions, but also to share experiences and to meet other social or commercial needs (Jiang and Zhu 2022a).

Therefore, in order to monitor mental health, it becomes necessary to first differentiate between the posts that genuinely seek support from the community and those that do not. Some examples of this ambiguity are shown in Figure 2. Note that the last example '[New] Week 22, can (you help) tell the gender?' was a community request. However, in Chinese culture, discussions about gender are considered a social 'talking point' rather than being directly related to health (Loo et al. 2009). Consequently, this example was deemed non-support-seeking. As a dataset for training the classifiers, 3000 posts were randomly selected and manually labelled as either support-seeking or non-support-seeking. Out of the 3000 posts, there were 1143

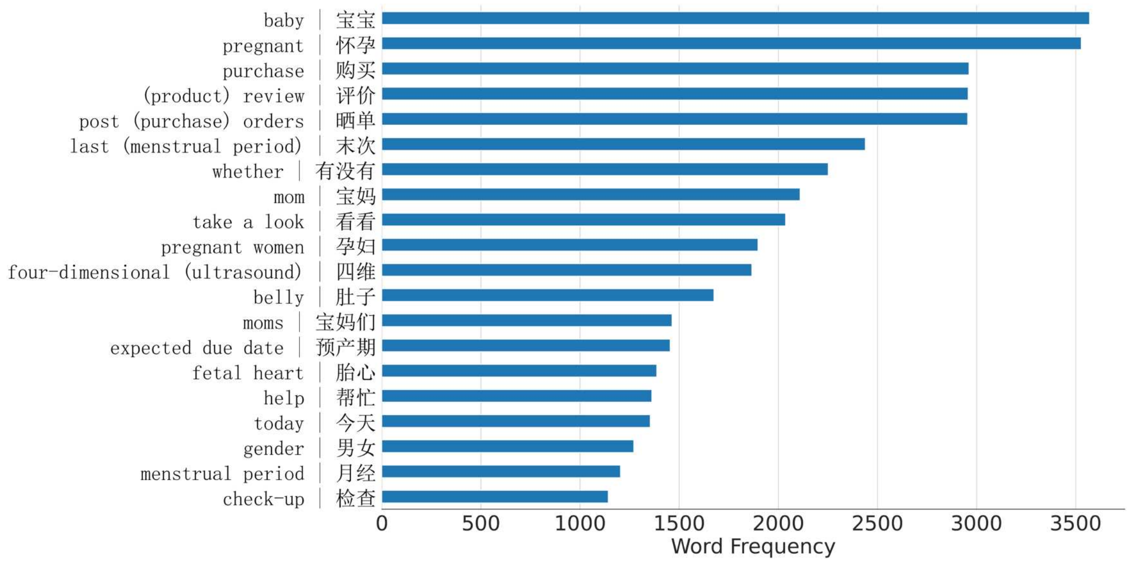

Figure 1. Top 20 most frequent words in user posts.

(38.1%) labelled as support-seeking and 1857 (61.9%) as non-support-seeking. Two annotators were hired to label this random sample independently. Both of them were experienced users of this online maternity community. Given that the two categories are mutually exclusive, the straightforward Cohen's κ coefficient (Cohen 1960) was employed to assess the agreement between the two annotators. Notably, the agreement reached near-perfection, with $\kappa = 0.962$.

3.2. Hypotheses

The online maternity communities such as the one studied here are triggered by a specific event, pregnancy. Therefore, the 'vocabulary' (i.e. the set of words used in these discussion forums) is expected to be stable across the 'birth clubs'. In other words, there is a significant information similarity embedded in the posts between different 'birth clubs', and a classifier trained on the data from one 'birth club' can also perform well with the data from a different 'birth club' in terms of differentiating support-seeking posts from non-support-seeking ones. As a result, the first hypothesis is formulated around the information similarity.

> H 1 (information similarity). *The predictive performance remains high if a classifier trained on the post data from one 'birth club' is applied to the post data from a different 'birth club'.*

As mentioned above, the other important dimension of the labelled post data is the pregnancy timeline. Since each trimester of pregnancy has its well established milestones and common symptoms, the 'vocabulary' is expected to vary along the timeline, even though it is not expected to vary between 'birth clubs'. Therefore, the second hypothesis is formulated around the time sensitivity.

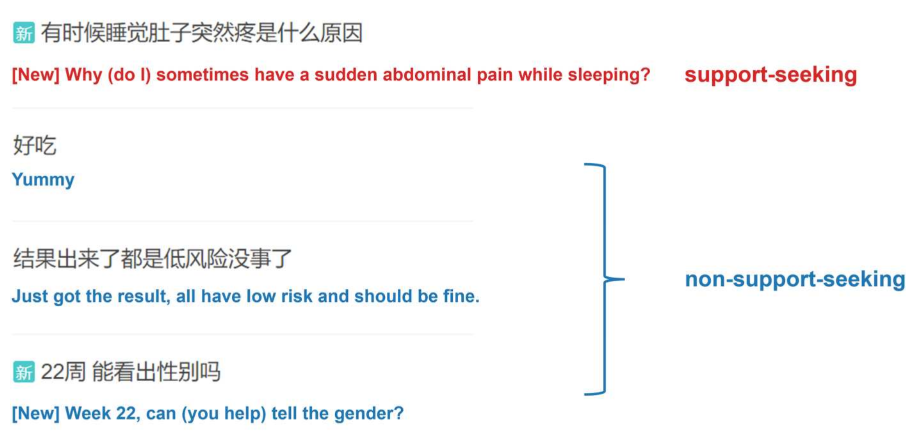

Figure 2. Examples of support-seeking and non-support-seeking posts.

H 2 (time sensitivity). *The predictive performance drops if a classifier trained on the post data from one trimester is applied to the post data from a different trimester.*

3.3. Methods

There is a growing scholarly interest in employing NLP for the detection of mental illness (Zhang et al. 2022), and more recently in the domain of maternal health care (Banik 2023; Bartal et al. 2023; De Choudhury, Counts, and Horvitz 2013; Luo et al. 2020). Give a relatively fixed 'vocabulary' around the event of pregnancy as discussed above, basic classification algorithms based on NLP are expected to achieve good predictive performance. Following standard NLP procedures, the posts were tokenised and the features were vectorised with the TF-IDF (term frequency-inverse document frequency) transformation. Then four basic classifiers were trained on the labelled data, including logistic regression, decision tree, random forest, as well as naive Bayes.

The naive Bayes (NB) classifier is a simple but highly effective algorithm with text data (Manning, Raghavan, and Schütze 2008). It is based on the probability of being in class c (say, support-seeking) given a document d:

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c)$$

$P(t_k|c)$ is the probability that feature t_k appearing in a document belonging to class c. Note that document d has n_d features.

The class that a given document d is assigned to is determined by the maximum a posteriori (MAP) class c_{map}, which is defined as:

$$\begin{aligned} c_{map} &= argmax_{c \in \mathbb{C}} \hat{P}(c|d) \\ &= argmax_{c \in \mathbb{C}} \hat{P}(c) \prod_{1 \leq k \leq n_d} \hat{P}(t_k|c) \end{aligned}$$

$\hat{P}(c)$ can be simply estimated as the number of documents belonging to class c divided by the total number of documents in the training set, whereas:

$$\hat{P}(t_k|c) = \frac{N_{ct_k} + \alpha}{N_c + \alpha n}$$

N_{ct_k} is the fractional counting of TF-IDF feature t_k appears in a sample of class c in the training set and N_c is the fractional counting of all TF-IDF features for class c in the training set. n is the number of TF-IDF features for class c. Finally, $\alpha \geq 0$ is the smoothing prior that accounts for features not present in the training set and prevents zero probabilities in future computations (Manning, Raghavan, and Schütze 2008).

4. Results

4.1. Baseline results

First, some baseline results were obtained with the four classifiers. Given the relatively small and label-imbalanced sample, a nested cross-validation (CV) approach was employed to enhance the generalisability of the results. An outer 5-fold stratified cross-validation was implemented to partition the labelled sample. As a result, in each of the five rounds, the 3000 labelled posts were split into a 80% training set (i.e. 2400 posts) and a 20% test set (i.e. 600 posts). Throughout the training process for each classifier, an inner 10-fold cross-validation procedure was applied to tune the hyperparameters, with the AUC (area under the curve) score serving as the evaluation metric. For each classifier, the N-gram range was included as a hyperparameter and could choose its value from {(1, 1), (1, 2)} (i.e. unigrams only or both unigrams and bigrams). For logistic regression, the regularisation parameter C could choose its value from {0.1, 1, 10}. For decision tree, the maximum depth could choose its value from {*None*, 5, 10}. For random forest, the number of estimators could choose its value from {100, 200, 300}. Finally for naive Bayes, the above mentioned α could choose its value from {0.1, 0.5, 1}. Although a more extensive exploration of the hyperparameter space is conceivable, the current choices of parameter sets suffice for adequately testing the hypotheses and thoroughly evaluating the results.

Table 2 reports the best average AUC score (averaged over both the inner and outer CV rounds) determining the hyperparameters for each classifier in the second column. The best model of each classifier was then applied to the test set and the results are reported in Table 2 with the standard evaluation metrics accuracy, precision, recall, and F1-score (averaged over the 5 outer CV rounds). The NB classifier outperformed others in nearly all metrics, especially with the highest accuracy of 0.8340 as well as the highest F1-score of 0.7733.

4.2. Information similarity

For the rest of the analysis, only the NB classifier was chosen because it consistently outperformed others (similar results were obtained with other classifiers for the following analysis and these results are available

Table 2. Baseline results.

Classifier	AUC	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.8968	0.8270	0.8108	0.7134	0.7578
Decision Tree	0.7884	0.7950	0.7798	0.6467	0.7059
Random Forest	0.8817	0.8136	0.8436	0.6286	0.7197
Naive Bayes	0.9041	0.8340	0.8058	0.7461	0.7733

upon request). To test **H1** regarding the information similarity, the post data of the 'birth club' March 2018 was used as a training set whereas the post data of the 'birth club' May 2018 was used as a test set (i.e. the training strategy named as '1st predicting 3rd BC' in Table 3). Alternatively, the post data of both 'birth clubs' March 2018 and April 2018 was used as a training set when predicting the data of May 2018 (i.e. the training strategy named as '1st & 2nd predicting 3rd BC' in Table 3). It can be seen that the extrapolation across 'birth clubs' still offered a comparable predictive performance. Both new training strategies achieved over 80% accuracy and over 70% F1-score. Understandably, the training strategy '1st & 2nd predicting 3rd BC' produced higher scores than the training strategy '1st predicting 3rd BC' because the former used a larger training set. Note that its accuracy of 0.8419 was even higher than the baseline. As a robustness check, in the last two rows respectively, the post data of the 'birth club' May 2018 (or of both 'birth clubs' April 2018 and May 2018) was used as a training set whereas the post data of the 'birth club' March 2018 was used as a test set (i.e. the training strategy named as '3rd predicting 1st BC' or '2nd & 3rd predicting 1st BC'). The patterns of the evaluation metrics exhibited no qualitative alterations. Therefore, **H1** is supported by the data.

4.3. Time sensitivity

On the other hand, to test **H2** regarding the time sensitivity, the post data of the first trimester was used as a training set whereas the post data of the third trimester was used as a test set (i.e. the training strategy named as '1st predicting 3rd TR' in Table 4). Alternatively, the post data of the first two trimester was used as a training set when predicting the data of the last trimester (i.e. the training strategy named as '1st & 2nd predicting 3rd TR' in Table 4). It can be seen that the extrapolation across trimesters compromised the predictive performance. Both new training strategies produced worse results compared with the baseline. Especially the training strategy '1st predicting 3rd TR' scored below 80% for accuracy and below 70% for F1-score. Although the training strategy '1st & 2nd predicting 3rd TR' improved the result with a larger training set (e.g. with the accuracy of 0.8166), its F1-score of 0.7131 was still far below the baseline. Furthermore, the findings were validated through a robustness check involving two supplementary experiments referred to as '3rd predicting 1st TR' and '2nd & 3rd predicting 1st TR' (i.e. the last two rows in Table 4). Therefore, **H2** is supported by the data.

Table 3. 'Birth clubs' extrapolation results.

Classifier	AUC	Accuracy	Precision	Recall	F1-Score
NB (baseline)	0.9041	0.8340	0.8058	0.7461	0.7733
NB (1st predicting 3rd BC)	0.8904	0.8167	0.7786	0.6785	0.7251
NB (1st & 2nd predicting 3rd BC)	0.9033	0.8419	0.7993	0.7428	0.7700
NB (3rd predicting 1st BC)	0.8698	0.7987	0.7603	0.7286	0.7441
NB (2nd & 3rd predicting 1st BC)	0.8909	0.8272	0.7898	0.7766	0.7832

Table 4. Trimester extrapolation results.

Classifier	AUC	Accuracy	Precision	Recall	F1-Score
NB (baseline)	0.9041	0.8340	0.8058	0.7461	0.7733
NB (1st predicting 3rd TR)	0.8591	0.7860	0.7490	0.5187	0.6130
NB (1st & 2nd predicting 3rd TR)	0.8879	0.8166	0.7291	0.6979	0.7131
NB (3rd predicting 1st TR)	0.9018	0.7249	0.6982	0.6205	0.6570
NB (2nd & 3rd predicting 1st TR)	0.9121	0.7780	0.7488	0.7182	0.7332

4.4. A monitoring case study

The previous analysis confirmed the properties of information similarity and time sensitivity of this online maternity community's user posts. The implication is that it is feasible to differentiate between support-seeking posts and non-support-seeking ones with relatively simple machine learning algorithms and small training sets. To demonstrate how this classification can be useful for monitoring maternal mental health, a case study was conducted with all 52,558 posts collected.

Figure 3 shows the number of posts during the pregnancy timeline arranged by weeks. The distribution was uneven as the volume of posts started off at a minimum number but quickly climbed up to over 2500 in Week 8. It then stayed between 1000 and 1500 for most of the second trimester, before significantly peaking again in Week 37. For an easier interpretation, Figure 3 also shows the most frequent words associated with the peaks at different points of time. For example, the most frequent word in Week 8 was 'last (menstrual period)', indicating that the users were actively trying to confirm their pregnancy in this week. Other frequently discussed topics such as 'NT (nuchal translucency)', 'baby movement', 'four-dimensional (ultrasound)', and 'baby' along the pregnancy timeline were discovered in the same fashion. Early intervention can be advised given the patterns discovered in Figure 3. For example, before Week 8, the information on pregnancy tests and miscarriage care can be shared and highlighted by 'birth club' moderators.

Then the best-performing baseline NB classifier was applied to all 52,558 posts and each post was classified

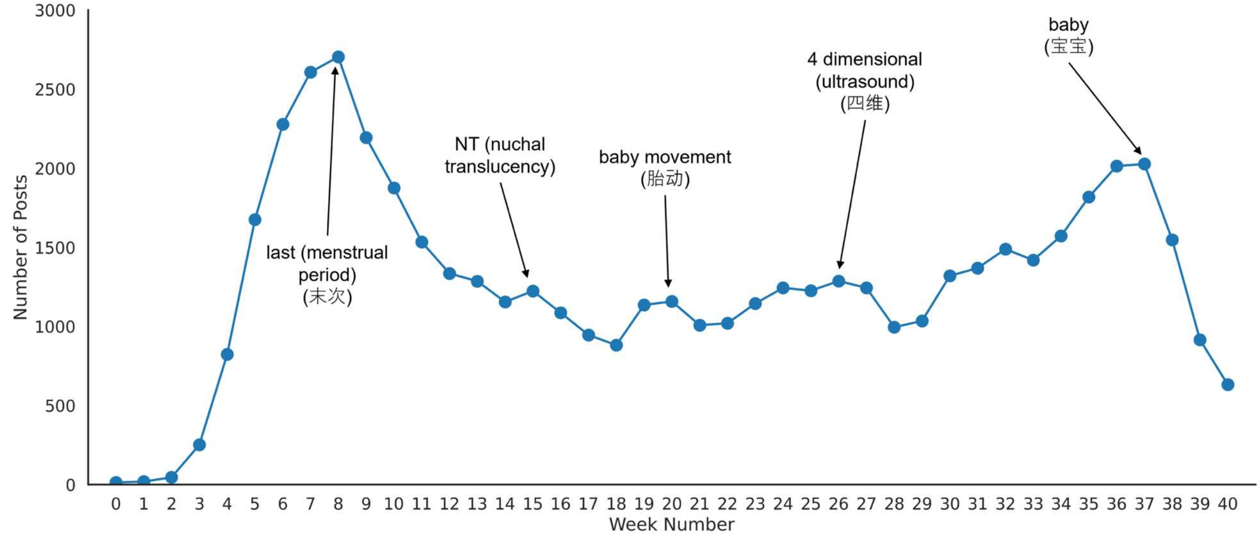

Figure 3. Number of posts during pregnancy.

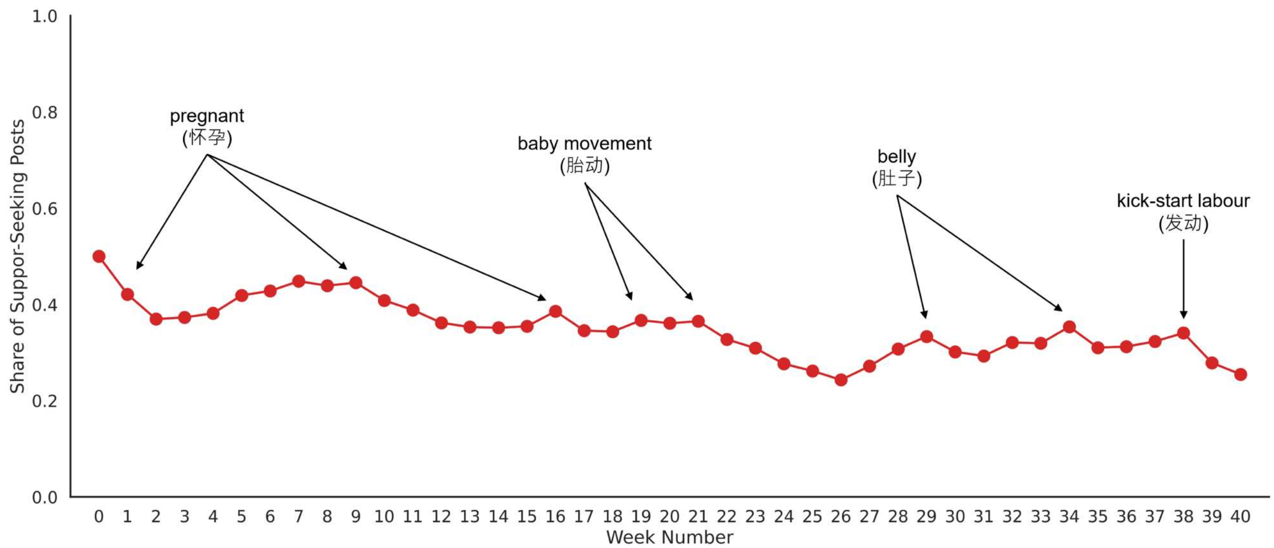

Figure 4. Share of support-seeking posts during pregnancy.

as either support-seeking or non-support-seeking. At each point of time, the proportion of support-seeking posts can be calculated as the number of support-seeking posts divided by the number of total posts. Because support-seeking posts often stem from mentally distressing concerns, such a proportion during the pregnancy can be monitored as a strong indicator for the maternal users' mental health status.

Unlike the uneven distribution observed in Figure 3, the share of support-seeking posts throughout the pregnancy timeline appeared to be stable, with a decreasing trend over the weeks. Therefore, the users collectively showed a moderate level of mental stress (around 30% as measured by the share of support-seeking posts) during their pregnancy and the level decreased gradually towards the end. The most frequent words in the support-seeking posts are also shown for some weeks in Figure 4. It can be seen that the users were initially more concerned with 'pregnant', and later with 'baby movement', 'belly', and 'kick-start labour'. If the patterns observed in Figure 4 are robust, any anomalies can be easily detected when significant deviations from the patterns occur (say, a sudden increase of the share of support-seeking posts to 60%).

5. Discussion and conclusion

Digital forms of social support offer a potential solution to addressing widespread maternal mental health concerns, thanks to advancements in digital technologies.

Consequently, there has been a global surge in popularity for maternity support web applications in recent years. Typically these platforms create discussion forums, often named as 'birth clubs', and make them available to users based on their estimated due months, facilitating the exchange of personal experiences and knowledge related to maternal care during pregnancy. Users within these peer groups can interact by submitting or responding to posts. It is important to note that a post does not always seek support but can also involve sharing personal experiences or meeting other social or commercial needs.

Therefore, a crucial initial step in monitoring maternal mental health on such digital platforms involves distinguishing between support-seeking and non-support-seeking posts. To achieve this, the present study employed NLP techniques on the data collected from one of the largest Chinese maternity and parenting web applications. The dataset consisted of 52,558 posts, with a subset of 3000 posts randomly selected and labelled as either support-seeking or non-support-seeking.

Four basic machine learning algorithms, logistic regression, decision tree, random forests, and naive Bayes (NB), were trained for post classification with the labelled post data. The NB classifiers perform the best in terms of standard evaluation metrics. Different training strategies were then implemented to validate the properties of information similarity and time sensitivity within the post data. Regarding the information similarity, a classifier trained with the post data from one 'birth club' still performed well with the post data from another 'birth club'. It implies that all 'birth clubs' discussed a relatively fixed set of topics. After all, this online maternity community is centred around a specific event, pregnancy. However, regarding the time sensitivity, a classifier trained with the post data from one trimester performed less effectively when applied to the post data from another trimester. This is due to the fact that each trimester of pregnancy comes with its own well established milestones and common symptoms.

To illustrate the practical application of this classification, a concrete case study encompassing all 52,558 posts was conducted to showcase how it can aid in monitoring users' mental health status, and how early intervention and anomaly detection can be implemented. The findings suggest the feasibility of utilising relatively straightforward machine learning algorithms and small training sets for monitoring maternal mental health. Looking ahead, this study could extend its scope to encompass other online maternity communities in other regions and languages, as well as to potentially explore individual-level mental health monitoring.

In the broader field of online mental health monitoring, the integration of advanced deep learning models, such as BERT (Devlin et al. 2019) and MentalBERT (Ji et al. 2022), presents a promising avenue for future exploration. By leveraging these models' context-aware embeddings, a more nuanced understanding of users' mental states becomes possible. Furthermore, extending these models to conversational agents, including large language models (LLMs) such as ChatGPT, has the potential to facilitate real-time analysis and proactive mental health support in online interactions. However, as we delve into these advancements, it is crucial to address ethical considerations and to uphold the importance of user privacy (Cronin et al. 2021; Martinez-Martin et al. 2020).

Acknowledgments

The author is grateful for the insightful comments received from the participants of the Inaugural International Digital Mental Health and Wellbeing Conference held at the Ulster University in June 2023. The usual disclaimers apply.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Adams, P. J.. 2008. *Fragmented Intimacy: Addiction in a Social World*. New York: Springer.

Alderdice, F., J. McNeill, and F. Lynn. 2013. "A Systematic Review of Systematic Reviews of Interventions to Improve Maternal Mental Health and Well-Being." *Midwifery* 29 (4): 389–399. https://doi.org/10.1016/j.midw.2012.05.010.

Banik, D.. 2023. "Evaluation of Maternal Patient Experience During COVID-19 Using Natural Language Processing." PhD thesis, The University of Texas at El Paso.

Bartal, A., K. M. Jagodnik, S. J. Chan, M. S. Babu, and S. Dekel. 2023. "Identifying Women with Postdelivery Posttraumatic Stress Disorder Using Natural Language Processing of Personal Childbirth Narratives." *American Journal of Obstetrics & Gynecology MFM* 5 (3): 100834. https://doi.org/10.1016/j.ajogmf.2022.100834.

Bedaso, A., J. Adams, W. Peng, and D. Sibbritt. 2021. "The Relationship Between Social Support and Mental Health Problems During Pregnancy: A Systematic Review and Meta-Analysis." *Reproductive Health* 18 (1): 1–23. https://doi.org/10.1186/s12978-020-01058-8.

Bennett, K., A. J. Bennett, and K. M. Griffiths. 2010. "Security Considerations for E-Mental Health Interventions." *Journal of Medical Internet Research* 12 (5): e1468. https://doi.org/10.2196/jmir.1468.

Broadbent, R., and T. Papadopoulos. 2013. "Bridging the Digital Divide—An Australian Story." *Behaviour & Information Technology* 32 (1): 4–13. https://doi.org/10.1080/0144929X.2011.572186.

Brown, M., N. O'Neill, H. van Woerden, P. Eslambolchilar, M. Jones, and A. John. 2016. "Gamification and Adherence to Web-Based Mental Health Interventions: A Systematic Review." *JMIR Mental Health*3 (3): e5710. https://doi.org/10.2196/mental.5710.

Chivers, B. R., R. M. Garad, J. A. Boyle, H. Skouteris, H. J. Teede, and C. L. Harrison. 2020. "Perinatal Distress During COVID-19: Thematic Analysis of An Online Parenting Forum." *Journal of Medical Internet Research* 22 (9): e22002. https://doi.org/10.2196/22002.

Cohen, J.. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37–46. https://doi.org/10.1177/001316446002000104.

Coppersmith, G., M. Dredze, and C. Harman. 2014. "Quantifying Mental Health Signals in Twitter." In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 51–60. Baltimore, Maryland, USA: Association for Computational Linguistics.

Cronin, B., N. Perra, L. E. C. Rocha, Z. Zhu, F. Pallotti, S. Gorgoni, G. Conaldi, and R. De Vita. 2021. "Ethical Implications of Network Data in Business and Management Settings." *Social Networks* 67:29–40. https://doi.org/10.1016/j.socnet.2020.09.001.

D'Alfonso, S.. 2020. "Ai in Mental Health." *Current Opinion in Psychology* 36:112–117. https://doi.org/10.1016/j.copsyc.2020.04.005.

De Choudhury, M., S. Counts, and E. Horvitz. 2013. "Predicting Postpartum Changes in Emotion and Behavior via Social Media." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3267–3276. New York, NY, USA: Association for Computing Machinery.

Denton, L. K., C. E. Creeley, B. Stavola, K. Hall, and B. D. Foltz. 2020. "An Analysis of Online Pregnancy Message Boards: Mother-to-Mother Advice on Medication Use." *Women and Birth* 33 (1): e48–e58. https://doi.org/10.1016/j.wombi.2018.12.003.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers), 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Di Cara, N. H., V. Maggio, O. S. Davis, and C. M. Haworth. 2023. "Methodologies for Monitoring Mental Health on Twitter: Systematic Review." *Journal of Medical Internet Research* 25:e42734. https://doi.org/10.2196/42734.

Drentea, P., and J. L. Moren-Cross. 2005. "Social Capital and Social Support on the Web: The Case of An Internet Mother Site." *Sociology of Health & Illness* 27 (7): 920–943. https://doi.org/10.1111/shil.2005.27.issue-7.

Drydakis, N.. 2021. "Mobile Applications Aiming to Facilitate Immigrants' Societal Integration and Overall Level of Integration, Health and Mental Health. Does Artificial Intelligence Enhance Outcomes?." *Computers in Human Behavior* 117:106661. https://doi.org/10.1016/j.chb.2020.106661.

Figueiredo, B., and A. Conde. 2011. "Anxiety and Depression in Women and Men From Early Pregnancy to 3-months Postpartum." *Archives of Women's Mental Health* 14 (3): 247–255. https://doi.org/10.1007/s00737-011-0217-3.

Gleeson, D. M., A. Craswell, and C. M. Jones. 2019. "Women's Use of Social Networking Sites Related to Childbearing: An Integrative Review." *Women and Birth* 32 (4): 294–302. https://doi.org/10.1016/j.wombi.2018.10.010.

Gui, X., Y. Chen, Y. Kou, K. Pine, and Y. Chen. 2017. "Investigating Support Seeking From Peers for Pregnancy in Online Health Communities." *Proceedings of the ACM on Human-Computer Interaction*1 (CSCW): 1–19. https://doi.org/10.1145/3134685.

Hermann, A., E. M. Fitelson, and V. Bergink. 2021. "Meeting Maternal Mental Health Needs During the COVID-19 Pandemic." *JAMA Psychiatry* 78 (2): 123–124. https://doi.org/10.1001/jamapsychiatry.2020.1947.

Hinduja, S., M. Afrin, S. Mistry, and A. Krishna. 2022. "Machine Learning-Based Proactive Social-Sensor Service for Mental Health Monitoring Using Twitter Data." *International Journal of Information Management Data Insights* 2 (2): 100113. https://doi.org/10.1016/j.jjimei.2022.100113.

Ji, S., T. Zhang, L. Ansari, J. Fu, P. Tiwari, and E. Cambria. 2022. "Mentalbert: Publicly Available Pretrained Language Models for Mental Healthcare." In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 7184–7190. Marseille, France: European Language Resources Association.

Jiang, L., and Z. Zhu. 2022a. "Information Exchange and Multiple Peer Groups: A Natural Experiment in An Online Community." *Journal of Economic Behavior & Organization* 203:543–562. https://doi.org/10.1016/j.jebo.2022.09.019.

Jiang, L., and Z. Zhu. 2022b. "Maternal Mental Health and Social Support From Online Communities During Pregnancy." *Health & Social Care in the Community* 30 (6): e6332–e6344.

Kim, J., J. Lee, E. Park, and J. Han. 2020. "A Deep Learning Model for Detecting Mental Illness From User Content on Social Media." *Scientific Reports* 10 (1): 11846. https://doi.org/10.1038/s41598-020-68764-y.

Lee, A. M., S. K. Lam, S. M. S. M. Lau, C. S. Y. Chong, H. W. Chui, and D. Y. T. Fong. 2007. "Prevalence, Course, and Risk Factors for Antenatal Anxiety and Depression." *Obstetrics & Gynecology*110 (5): 1102–1112. https://doi.org/10.1097/01.AOG.0000287065.59491.70.

Loo, K. K., X. Luo, H. Su, A. Presson, and Y. Li. 2009. "Dreams of Tigers and Flowers: Child Gender Predictions and Preference in An Urban Mainland Chinese Sample During Pregnancy." *Women & Health* 49 (1): 50–65. https://doi.org/10.1080/03630240802694673.

Luo, X., H. Ding, M. Tang, P. Gandhi, Z. Zhang, and Z. He. 2020. "Attention Mechanism with Bert for Content Annotation and Categorization of Pregnancy-Related Questions on a Community Q&A Site." In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1077–1081. Seoul, South Korea: IEEE.

Manning, C., P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Martinez-Martin, N., I. Dasgupta, A. Carter, J. A. Chandler, P. Kellmeyer, K. Kreitmair, A. Weiss, and L. Y. Cabrera. 2020. "Ethics of Digital Mental Health During COVID-19: Crisis and Opportunities." *JMIR Mental Health* 7 (12): e23776. https://doi.org/10.2196/23776.

McClellan, C., M. M. Ali, R. Mutter, L. Kroutil, and J. Landwehr. 2017. "Using Social Media to Monitor Mental Health Discussions- Evidence From Twitter." *Journal of the American Medical Informatics Association* 24 (3): 496–502. https://doi.org/10.1093/jamia/ocw133.

Naslund, J. A., K. A. Aschbrenner, and S. J. Bartels. 2016. "Wearable Devices and Smartphones for Activity Tracking Among People with Serious Mental Illness." *Mental Health and Physical Activity*10:10–17. https://doi.org/10.1016/j.mhpa.2016.02.001.

Onyeaka, H., J. Firth, R. C. Kessler, K. Lovell, and J. Torous. 2021. "Use of Smartphones, Mobile Apps and Wearables for Health Promotion by People with Anxiety Or Depression: An Analysis of a Nationally Representative Survey Data." *Psychiatry Research* 304:114120. https://doi.org/10.1016/j.psychres.2021.114120.

Orr, S. T.. 2004. "Social Support and Pregnancy Outcome: A Review of the Literature." *Clinical Obstetrics and Gynecology* 47 (4): 842–855. https://doi.org/10.1097/01.grf.0000141451.68933.9f.

Orr, S. T., and C. A. Miller. 1997. "Unintended Pregnancy and the Psychosocial Well-Being of Pregnant Women." *Women's Health Issues* 7 (1): 38–46. https://doi.org/10.1016/S1049-3867(96)00017-5.

Parker, L., V. Halter, T. Karliychuk, and Q. Grundy. 2019. "How Private is Your Mental Health App Data? An Empirical Study of Mental Health App Privacy Policies and Practices." *International Journal of Law and Psychiatry* 64:198–204. https://doi.org/10.1016/j.ijlp.2019.04.002.

Rahman, A., P. J. Surkan, C. E. Cayetano, P. Rwagatare, and K. E. Dickson. 2013. "Grand Challenges: Integrating Maternal Mental Health Into Maternal and Child Health Programmes." *PLoS Medicine* 10 (5): e1001442. https://doi.org/10.1371/journal.pmed.1001442.

Ross, L. E., L. M. McLean, and C. Psych. 2006. "Anxiety Disorders During Pregnancy and the Postpartum Period: A Systematic Review." *Depression* 6 (9): 1–14.

Sardi, L., A. Idri, and J. L. Fernández-Alemán. 2017. "A Systematic Review of Gamification in E-Health." *Journal of Biomedical Informatics* 71:31–48. https://doi.org/10.1016/j.jbi.2017.05.011.

Silva-Jose, C., T. S. Nagpal, J. Coterón, R. Barakat, and M. F. Mottola. 2022. "The 'New Normal' includes Online Prenatal Exercise: Exploring Pregnant Women's Experiences During the Pandemic and the Role of Virtual Group Fitness on Maternal Mental Health." *BMC Pregnancy and Childbirth* 22 (1): 1–11. https://doi.org/10.1186/s12884-022-04587-1.

Tommasel, A., A. Diaz-Pace, D. Godoy, and J. M. Rodriguez. 2022. "Tracking the Evolution of Crisis Processes and Mental Health on Social Media During the COVID-19 Pandemic." *Behaviour & Information Technology* 41 (16): 3450–3469. https://doi.org/10.1080/0144929X.2021.1998630.

Van Der Krieke, L., L. Wunderink, A. C. Emerencia, P. De Jonge, and S. Sytema. 2014. "E–Mental Health Self-Management for Psychotic Disorders: State of the Art and Future Perspectives." *Psychiatric Services* 65 (1): 33–49. https://doi.org/10.1176/appi.ps.201300050.

Wachs, T. D., M. M. Black, and P. L. Engle. 2009. "Maternal Depression: A Global Threat to Children's Health, Development, and Behavior and to Human Rights." *Child Development Perspectives* 3 (1): 51–59. https://doi.org/10.1111/cdep.2009.3.issue-1.

Wexler, A., A. Davoudi, D. Weissenbacher, R. Choi, K. O'Connor, H. Cummings, and G. Gonzalez-Hernandez. 2020. "Pregnancy and Health in the Age of the Internet: A Content Analysis of Online "Birth Club" Forums." *PLoS ONE* 15 (4): e0230947. https://doi.org/10.1371/journal.pone.0230947.

Zhang, T., A. M. Schoene, S. Ji, and S. Ananiadou. 2022. "Natural Language Processing Applied to Mental Illness Detection: A Narrative Review." *NPJ Digital Medicine* 5 (1): 46. https://doi.org/10.1038/s41746-022-00589-7.

Zhou, J., K. L. Havens, C. P. Starnes, T. A. Pickering, N. H. Brito, C. L. Hendrix, M. E. Thomason, T. C. Vatalaro, and B. A. Smith. 2021. "Changes in Social Support of Pregnant and Postnatal Mothers During the COVID-19 Pandemic." *Midwifery* 103:103162. https://doi.org/10.1016/j.midw.2021.103162.