How do Employees Engage Within an Online Wellbeing Platform? Insights From Association Rule Mining Analysis

Gillian Cameron[1,2,*], Maurice Mulvenna[2], Raymond Bond[2], Edel Ennis[3], Siobhan O'Neill[3], David Cameron[1] and Alex Bunting[1]

[1] *Inspire, Belfast, Northern Ireland*

[2] *School of Computing, Faculty of Computing Engineering & the Built Environment, Ulster University, Belfast, Northern Ireland*

[3] *School of Psychology, Faculty of Life and Health Sciences, Ulster University, Coleraine, Northern Ireland*

Abstract

The increasing availability of mental health data presents both opportunities and challenges, particularly due to the unstructured and noisy nature of such data. Data mining—an analytical approach for extracting knowledge from large datasets—is becoming increasingly prevalent in the fields of medicine and mental health. By employing data mining techniques, the insights can help inform the development of enhanced digital tools for mental health and fostering a more personalised user experience. One notable method within this domain is association rule mining, which identifies frequent relationships between items in a dataset. This study aims to apply association rule mining to a dataset generated by users of a digital employee wellbeing platform, focusing on the relationships between various tools and resources utilised on the platform.

The Inspire Support Hub is a digital employee wellbeing platform featuring tools such as a mood tracker, a chatbot for self-assessments, and psychoeducational resources. User interactions with the platform are logged as anonymous events, including clicks, mood entries, and self-assessment results, each associated with a unique user ID and timestamp. Upon registration, users enter a company pin and their sector is recorded. From February 2019 to April 2023, 11,583 users engaged with the platform over 16,657 sessions. The analysis was conducted using R Studio, employing the dplyr and tidyverse packages for data cleaning and wrangling, along with ggplot2 for visualisation. The event logs were transformed into transaction data for association rule mining using the arules package. The Apriori algorithm was applied with a minimum support threshold of 0.05 and a confidence level of 0.8, ensuring that only rules with at least 80% accuracy were included.

Applying association rule mining on the employee wellbeing platform dataset revealed distinct sets of co-associations, with significant emphasis on the chatbot and mood tracker. This is predictable, considering that the iHelpr chatbot, anxiety and stress self-assessments, and mood tracker are the platform's most frequently used components. The association rules derived from this analysis can offer valuable insights into the user journey, such as discovering frequent usage patterns, and based upon this the platform could recommend personalised features and content tailored to each user's preferences and behaviour.

Keywords
Machine learning, employee wellbeing, employee mental health, association rule mining

1. Introduction

This paper discusses the application of the machine learning technique, association rule mining to an event log dataset produced by a digital employee wellbeing platform. The platform consists of a

AICS'24: 32nd Irish Conference on Artificial Intelligence and Cognitive Science, December 09—10, 2024, Dublin, Ireland
* Corresponding author.
cameron-g2@ulster.ac.uk (G. Cameron)
0009-0002-4921-0482 (G. Cameron); 0000-0002-1554-0785 (M. Mulvenna); 0000-0002-9677-0725 (E. Ennis); 0000-0002-8786-2118 (S. O'Neill); 0000-0002-1078-2232 (R. Bond); 0009-0008-3933-8072 (D. Cameron); 0009-0000-2327-2541 (A. Bunting)

chatbot to conduct self-assessment, psychoeducational resources, and trackers such as a mood and sleep tracker. This paper will address the background on digital tools within the workplace that are

designed to support mental health and wellbeing in the workplace, and how machine learning techniques have been applied to the mental health domain.

Results from applying association rule mining to a dataset of over 11,000 employees will be discussed, with limitations, practical implications and directions for future research.

1.1. Digital tools to support mental health in the workplace

Mental health in the workplace has gained significant attention in recent years, due to the significant impact on employee wellbeing, productivity and overall organisational performance. It is estimated that there were 875,000 cases of work-related stress, anxiety or depression in Great Britain in 2022/23, rising higher than the pre-pandemic level [1].
Research has shown that initiatives such as Employee Assistance Programmes (EAPs) can improve employee outcomes, particularly presenteeism and functioning [2]. Employee mental health symptoms and conditions, including stress, anxiety, depression, burnout, and psychological wellbeing, have also been found to respond well to digital interventions [3][4]. Due to these innovations in mental health, data on mental health in the workplace is becoming more widely available, but because it is often noisy or unstructured, processing it can be difficult.

1.2. Overview of Association Rule Mining

Data mining, which involves extracting knowledge from large datasets, is gaining popularity in medicine and mental health. Association rule mining is a prominent technique in this field, discovering frequent relationships between items in a dataset.
Association rule mining is a rule-based unsupervised machine learning technique, for discovering frequent patterns, associations and relationships in large datasets. Association rule mining has been applied to many domains, with the most known application detecting regularities between items in a supermarket, introduced by Agrawal, Imieliński and Swami in 1993 [5]. By applying association rule mining, we can detect rules within a dataset. For example, the rule {bread, milk} => {butter} found in supermarket datasets would indicate that if a customer buys both bread and milk, they are also likely to buy butter. These insights can be utilised to help with item placement, promotions or product recommendations to customers.

Association rule mining is now frequently being applied to other domains, such as mental health, including analysing the self-reported mental health symptoms of college students [6], exploring ADHD comorbidity [7], understanding why people call crisis helplines [8] and most relevant to this study, understanding user's interactions with a digital mental health application [9].
In the context of this study, association rule mining has been applied to event logs, produced by a digital employee wellbeing platform, in order to identify patterns in employee wellbeing behaviours.

In partnership with Inspire, a social enterprise that provides Employee Assistance Programmes (EAPs), focuses on analysing how a digital mental health intervention is used in the real world by employees across various sectors, by using association rule mining. The research questions for this study are:
RQ1: What are the most prevalent events logged on a digital employee wellbeing platform?
RQ2: Can association rule mining find related events within the digital employee wellbeing platform dataset?

2. Methodology

2.1. The Inspire Support Hub

The Inspire Support Hub is an employee wellbeing platform that can be accessed on any device by employees who utilise Inspire’s EAP. It has been developed using PHP, MySQL, HTML5, JavaScript, and CSS, and hosts a range of features, including e-learning programmes, a wellbeing library, a mood and sleep tracker, gratitude diary and video library. There is also a “quick link” menu, with popular resources such as anxiety, stress, and depression. The modular e-learning programmes are based on Cognitive Behavioural Therapy principles, and are on the following topics: stress, anxiety, depression, alcohol and self-esteem. An example screen of the stress e-learning programme is shown in Figure 1.

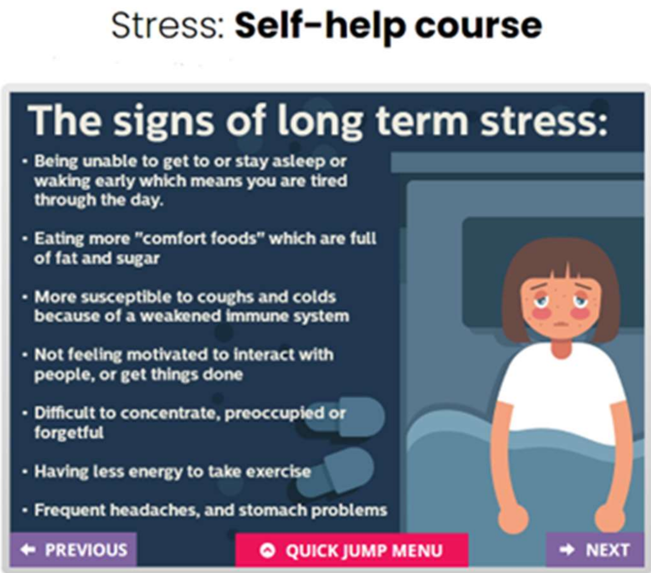

Figure 1 - stress e-learning programme

The employee can also access guided self-assessment on the following topics: stress, anxiety, alcohol, depression, sleep, and self-esteem through a chatbot called iHelpr, shown in Figure 2. The conversational script for iHelpr was developed by Inspire’s Clinical Lead [10]. iHelpr presents the user with validated screening instruments including the GAD7 for Anxiety [11], PHQ9 for Depression [12], Perceived Stress Scale for Stress [13], the Audit questionnaire for alcohol [14] the Sleep Condition Indicator for sleep [15] and a single item indicator for Self-Esteem. After completing a questionnaire, the user receives a personalised report from the chatbot with self-help recommendations, such starting an e-learning program, or, if their score is higher, they are routed to Inspire's in-person services or helpline.

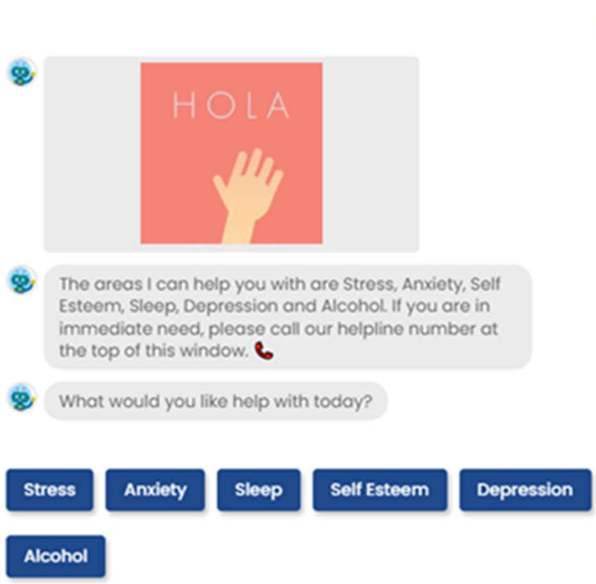

Figure 2 - iHelpr - Chatbot for Guided Self-Assessment

2.2. The Inspire Support Hub data

The event logs that were recorded on the Inspire Support Hub between February 2019 and April 2023 serve as the study's dataset. The dataset is made up of anonymous events, and each user has a unique ID. Each event is also timestamped to allow for the examination of occurrences over time. Clicks on pages and buttons within platform, sleep and mood logs, and self-assessment scores via the chatbot are all logged as events, and there are a possible 503 distinct events that could be accessed within the platform. When a user is onboarded to the employee wellbeing platform the sign up utilising a “company pin”. This is unique to their organisation who have implemented Inspire’s EAP services and allows for the identification of an industry/sector. Ulster University's Faculty of Computing, Engineering and the Built Environment ethics filter committee has approved this project and Inspire have conducted a data protection impact assessment.

2.3. Data Analysis

R studio and programming language was utilised for data wrangling, data cleaning and all analysis. Each user session was converted to a “transaction” to conduct the analysis. The arules package was installed for the association rule mining technique. The ggplot2 library was used for data visualisations, as well as arulesViz for visualising the rules.

2.4. Association Rule Mining

Introduced by Agrawal and Srikant in 1994 [16], the Apriori algorithm is a foundational technique within association rule mining, by identifying frequent item sets, then generating association rules based on specified confidence levels.

Using the Apriori algorithm associations between events accessed by users can be found during the user’s tenure on the employee wellbeing platform, where tenure is defined as the period of time from their first and last interactions within the data set.

In the context of this research a rule can be interpreted as:

If a user’s session contains event A, then event B is likely to be present in a later interaction with the digital employee wellbeing platform.

The Apriori algorithm was applied to the event log dataset, setting the minimum support to 0.05, meaning that events with a proportion of the total available events within the dataset of over 5 percent were included in the association rule mining. The confidence level was set to 0.8, meaning that the rule was only included in the output if it is correct 80% of the time. 74 rules were returned from a dataset of 508 distinct events and 16,657 user sessions.

3. Results

3.1. Descriptive statistics

Between February 2019 and the April 2023 several hundred client companies were set up on the platform, across 13 sectors, with 11,583 users, and 139,622 events logged over 16,657 user sessions. The platform was utilised primarily between the hours of 9am and 5pm on a weekday, with 80.47% of all interactions occurring during this time frame (Figure 3).

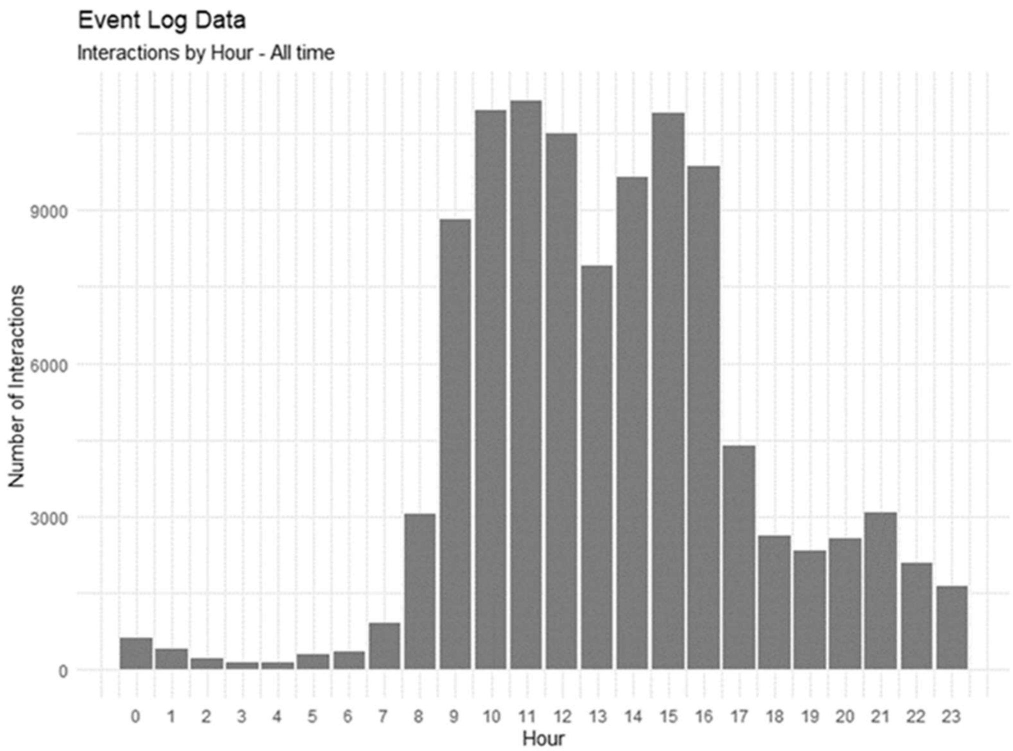

Figure 3 - Event Log frequency by Hour

3.2. Frequency of event usage

RQ1 was to find commonly used accessed by employees on a wellbeing platform. Figure 4 presents the most frequent events logged, and the top ten items are described below:

1) chatbot – guided self-assessment on anxiety, stress, depression, alcohol, sleep and self-esteem, described above in section 2.1.
2) expand menu - clicking on the burger menu when the platform is accessed on a mobile device
3) dashboard - "homepage" area of the platform, where users can see recently accessed items, data from their hub usage
4) self-help library - wellbeing literature on various mental health topics
5) online self-help - the e-learning modules described in section 2.1
6) take5search – a searchable database based on the take 5 ways to wellbeing

7) diary – the mood and sleep tracker area, where users can see their data presented back in graphs
8) quicklinkanxiety – most common topics (Anxiety, Depression, Stress) had easy to access "quick links" within the dashboard
9) howwecanhelp – an area on the dashboard detailing other support services, such as face to face counselling and a helpline number
10) moodtrack – the user's mood and sleep track logged to the database

When looking at specific mental health topics, anxiety and stress were the most frequently accessed, followed by information on looking after your mental health during the Covid-19 pandemic.

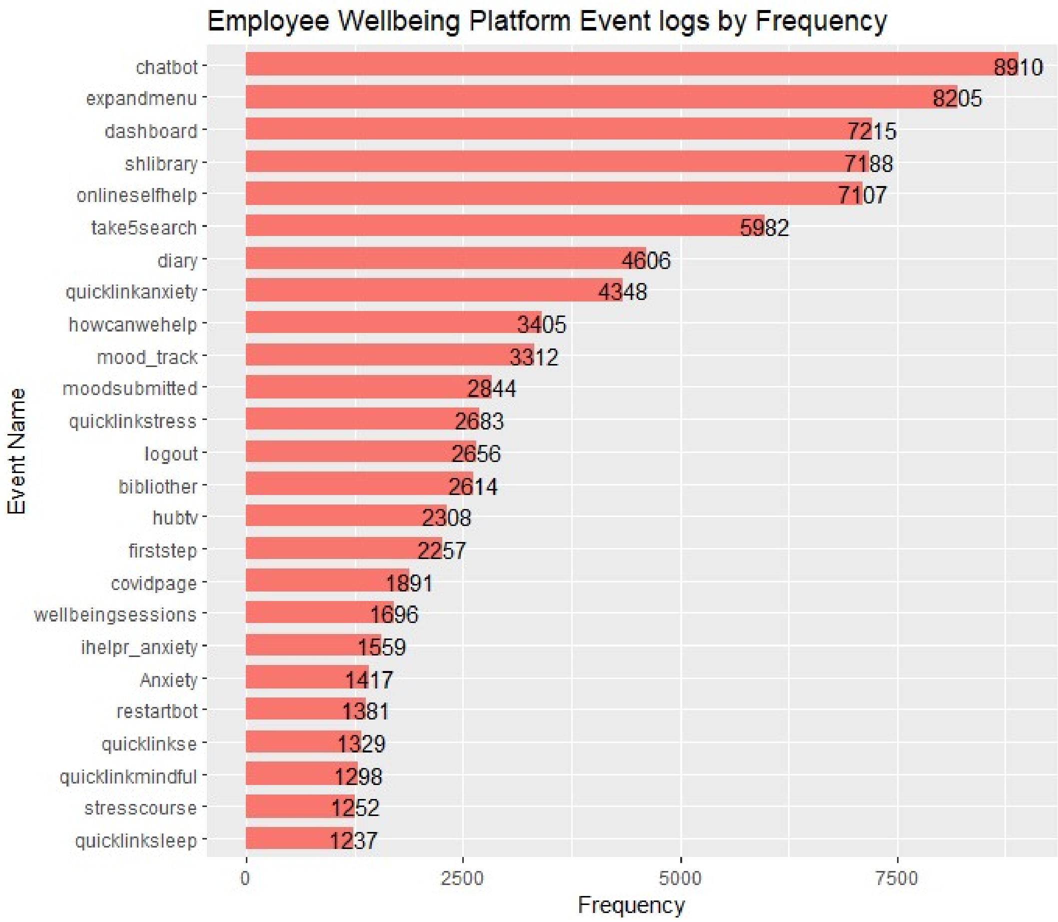

Figure 4 - Events log type by frequency

3.3. Key Findings from Association Rule Mining

In order to answer RQ2, the following results have been derived from applying association rule mining to a dataset containing event logs from an employee wellbeing platform.
A total of 74 rules were returned from a dataset containing 508 distinct events and 16,657 user sessions. The 74 rules are depicted in Figure 5, in terms of confidence, support and lift.

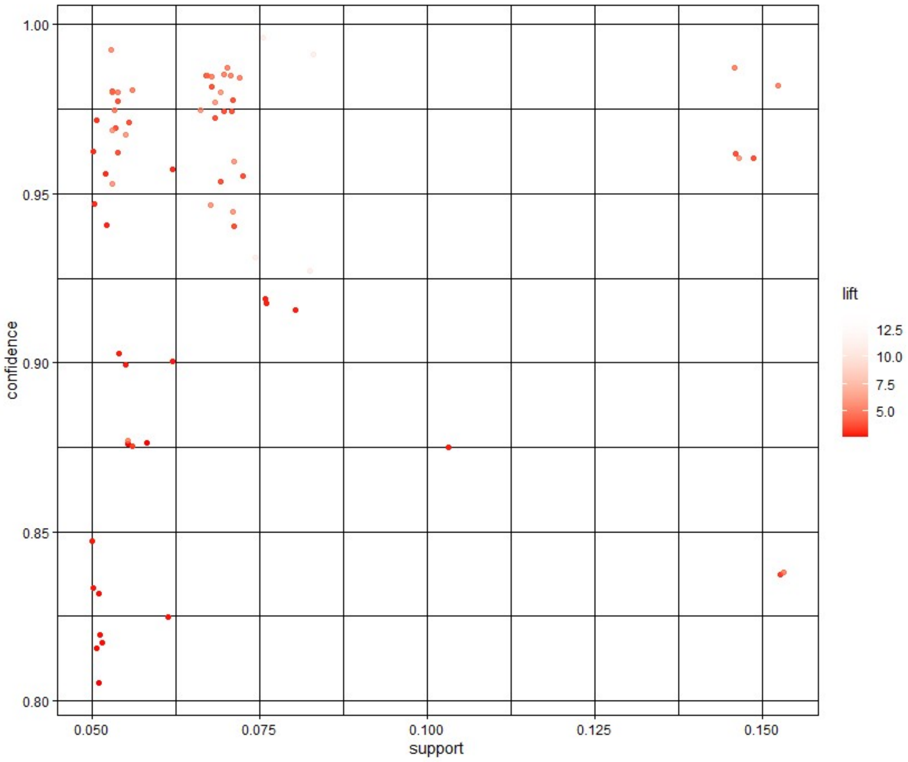

Figure 5 – A total of 74 rules from association rule mining, by confidence, support and lift

The top three rules will be ranked below by confidence, support and lift. The top 10 rules, sorted by confidence, are displayed below in Table 1.

Table 1 - Top 10 rules (sorted by confidence descending)

LHS	RHS	Support	Confidence	Lift	Count
{Anxiety,chatbot}	{ihelpr_anxiety}	0.08	0.99	11.26	1256
{Anxiety}	{ihelpr_anxiety}	0.08	0.99	11.24	1365
{quicklinkanxietyselfassessment}	{quicklinkanxiety}	0.05	0.99	5.82	868
{chatbot,Stress}	{ihelpr_stress}	0.05	0.99	14.32	908
{diary,moodsubmitted}	{mood_track}	0.15	0.99	5.41	2448
{Stress}	{ihelpr_stress}	0.06	0.99	14.30	1008
{moodsubmitted,take5search}	{mood_track}	0.07	0.99	5.41	1132
{moodsubmitted,onlineselfhelp}	{mood_track}	0.07	0.99	5.41	1196
{diary,moodsubmitted,take5search}	{mood_track}	0.07	0.99	5.41	1113
{diary,moodsubmitted,onlineselfhelp}	{mood_track}	0.07	0.99	5.41	1167

3.3.1. Confidence

In the context of this research, confidence is a measure of the strength of the association between two events, so if event A on the left-hand side is present, it is a strong predictor of the presence of the right-hand event.

When sorted by confidence descending, the leading rule is:

{Anxiety,chatbot} => {ihelpr_anxiety} (Confidence: 0.99)

When the user accesses the Anxiety information page, and the Chatbot, then they are likely to log an anxiety self-assessment.

When sorted by confidence descending, the second leading rule is:

{Anxiety} => {ihelpr_anxiety} (Confidence: 0.99)

When a user accesses the Anxiety information page, they are likely to log an anxiety self-assessment.

When sorted by confidence descending, the third leading rule is:

{qlanxietysa} => {quicklinkanxiety} (Confidence: 0.99)

When a user accesses the anxiety self-assessment from the quick link menu, they are likely to access other anxiety materials within the quick link menu.

3.3.2. Support

In the context of this research, support highlights how frequently the rule appears in the dataset. When sorted by support descending, the leading rule is:

{mood_track} => {diary} (Support: 0.15)

When a user logs a mood track, they are likely to then click diary, which is where they could view their mood and sleep logs across the month.

When sorted by support in descending order, the second leading rule is:

{moodsubmitted} => {mood_track} (Support: 0.15)

When a user submits their mood track, their mood track result is stored in the database.

When sorted by support descending, the third leading rule is:

{mood_track} => {moodsubmitted} (Support: 0.15)

When a user submits a mood track, they are likely to submit another mood track in a later interaction with the platform.

3.3.3. Lift

In the context of this research, lift tells us how much more likely events are to occur together, compared to happening independently. If the lift value of a rule is greater than 1, then this indicates that the events are logged together more often than what would be expected by chance.

When sorted by lift in descending order, the leading rule is:

{chatbot,Stress} => {ihelpr_stress} (Lift: 14.32)

When the user accesses the chatbot, and the page on Stress information, then they are likely to log a stress self-assessment.

When sorted by lift in descending order, the second leading rule is:

{chatbot,ihelpr_stress} => {Stress} (Lift: 14.31)

When a user accesses the chatbot, and they have submitted a stress self-assessment, they are likely to click on other materials on Stress.

When sorted by lift in descending order, the third leading rule is:

{ihelpr_stress} => {Stress} (Lift: 14.30)

When a user completes a stress self-assessment, they are likely to click on other materials on Stress.

Figure 6 is a visualisation of the association rule mining results, presented as a matrix-based plot, with the two key metrics of lift and support.

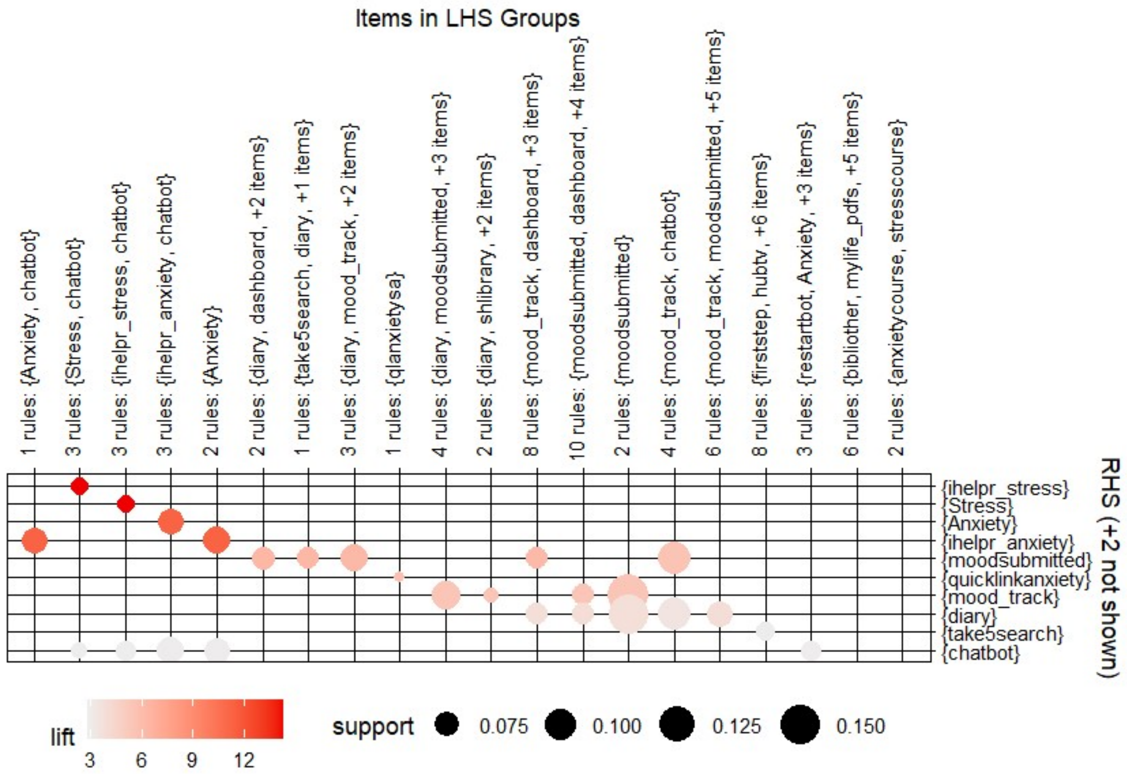

Figure 6 – Matrix-based visualisation of rules by life and support

4. Discussion

This study set out to discover insights and patterns in a dataset of event logs derived from an employee wellbeing platform, and address the following research questions:

RQ1: What are the most prevalent events logged on a digital employee wellbeing platform?
Frequently used features within the platform were interactive, allowing the users to input data, such as a self-assessment questionnaire, or tracking their mood and sleep. These features were utilised more often than psychoeducational materials such as the self-help library and e-learning programmes. These results indicate users may prefer to interact with features that give them feedback about their mental health, such as a summary of their self-assessment scores or mood tracks. This finding can be used to design digital mental health platforms, by incorporating more interactivity within other elements, such as the psychoeducational material.

RQ2: Can association rule mining find related events within the digital employee wellbeing platform dataset?

Clear sets of co-associations were found within the dataset, with the chatbot and mood tracker featuring heavily within the rules identified. This is to be expected, given that the most frequently utilised components of the platform include the iHelpr chatbot, particularly the anxiety and stress self-assessments, and the mood tracker. Association rule mining can often generate a very large number of rules, not all of which are meaningful or actionable, therefore it is important to set the minimum thresholds of support, lift and confidence relative to the dataset being analysed.

The platform architecture and design may significantly influence how a user interacts with the platform, and consequently the outcomes of association rule mining. Creating an intuitive and user-friendly platform can facilitate easier navigation between features, increasing the likelihood of co-associations being present. Furthermore, a cluttered or poorly designed menu or dashboard may hinder user engagement, leading to skewed data that underrepresents potential co-associations.

4.1. Implications for Employee Wellbeing Platforms

There could be potential to use the association rules to detect patterns indicative of declining wellbeing, for example certain combinations such as frequent negative mood tracks, or self-assessment scores, and sharp incline in searches for mental health topics. This could be developed into an automated system to remind employees of face-to-face services available, or to facilitate the automated deployment of a just-in-time interventions [17] or ecological momentary interventions [18].

These results offer practical insights when developing digital employee wellbeing platforms, by perhaps prompting developers to enhance commonly used components or integrate them more seamlessly. For example, placing frequently utilised features (in this case the mood diary and the chatbot) together to create a more natural pathway for the user to access these tools.

With association rule mining identifying frequent item sets, the platform could be adapted to recommend personalised features and content based on common usage pathways, for example if a user continuously logs lower moods or recorded sleep hours, the platform could recommend looking at guides on managing their mood and sleep.

4.2. Future Research Directions

Replicating this approach on subsets of the dataset, including content types such as videos, or written content, would be a good starting point to look for any trends in the information employees are seeking about their mental health. By filtering the dataset to include only items that are utilised for a duration exceeding a specific threshold, such as more than one minute, we may uncover more meaningful and robust co-associations within the dataset.

In conjunction with this, by enhancing the platforms capability to allow users to give feedback on the features in real time, would offer insights to utilise alongside the co-associations, to deliver personalised content by anticipating user preferences more precisely. Furthermore, creating a real-time user feedback mechanism could help to refine the UI continuously, ensuring it evolves according to actual user needs.

Further analysis could be conducted in the context of applying association rule mining to understand what rules are present within different user clusters. Further analysis on this dataset has been completed using k-means clustering to identify 3 user groups within the employee wellbeing platform; short-term, intermediate and long-term users. Understanding which features are highly utilised and used together within these clusters could inform strategies of improving user engagement and experience. Moreover, KModes, a clustering algorithm within data science, designed to organise similar data points into clusters according to their categorical features could be employed on the user event logs.

Further analysis of how rules may evolve over time, or within different sectors/industries with varying workloads and psychological stressors at work could be explored.

5. Conclusion

To summarise, we found clear sets of co-associations, which are arguably expected due to the nature of layout of the tools within the digital employee wellbeing platform. Features that allow the user to input their self-reported data, such as self-assessment questionnaires, or mood trackers are utilised more than psychoeducational materials. Based on these association rules, we can obtain greater insight into the user's journey and provide insights and recommendations for building better digital tools for mental health, with a more personalised user experience that could be driven by insights derived from AI techniques such as association rule mining.

5.1. Limitations of the Study

The platform only collects anonymised data, excluding the user's industry of employment and any demographic data. Users cannot be contacted for additional research, and demographic data cannot be used to examine how different age groups or genders use the digital employee wellbeing platforms.

Furthermore, as the platform does not collect emails, users cannot reset their password. This could lead to users not logging in at all or opening a new account, which would result in the assignment of a new anonymous ID that is unrelated to their previous account.

Acknowledgements

This work has been supported by the Industrial Fellowship programme by the Royal Commission for the Exhibition of 1851. This work would not have been possible without the ongoing support from Inspire.

References

[1] Health and Safety Executive (HSE), 2023. Work-related ill health and injury statistics 2022/23. Available at: https://www.hse.gov.uk/statistics/

[2] Joseph, B., Walker, A. and Fuller-Tyszkiewicz, M., 2018. Evaluating the effectiveness of employee assistance programmes: A systematic review. European Journal of Work and Organizational Psychology, 27(1), pp.1-15.

[3] Vertola, G., Marcello, A., Bottone, M., Sperandeo, R., Muzii, B., Scandurra, C. and Maldonato, N.M., 2022, September. Use and Effectiveness of Mobile Health Applications for Stress Management and Emotional Self-Regulation in Adult Workers: A Systematic Review. In 2022 13th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000081-000086). IEEE.

[4] Phillips, E.A., Gordeev, V.S. and Schreyögg, J., 2019. Effectiveness of occupational e-mental health interventions. Scandinavian journal of work, environment & health, 45(6), pp.560-576.

[5] Agrawal, R., Imieliński, T. and Swami, A., 1993, June. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (pp. 207-216).

[6] Zhao, L., Hao, F., Xu, T. and Dong, X., 2017. Positive and negative association rules mining for mental health analysis of college students. EURASIA Journal of Mathematics, Science and Technology Education, 13(8), pp.5577-5587.

[7] Tai, Y.M. and Chiu, H.W., 2009. Comorbidity study of ADHD: applying association rule mining (ARM) to National Health Insurance Database of Taiwan. International journal of medical informatics, 78(12), pp.e75-e83

[8] Turkington, R., Mulvenna, M.D., Bond, R.R., O'Neill, S., Potts, C., Armour, C., Ennis, E. and Millman, C., 2020. Why do people call crisis helplines? Identifying taxonomies of presenting reasons and discovering associations between these reasons. Health informatics journal, 26(4), pp.2597-2613.

[9] Alslaity, A., Chan, G., Wilson, R. and Orji, R., 2022, December. Toward Understanding Users' Interactions with a Mental Health App: An Association Rule Mining Approach. In International Conference on Pervasive Computing Technologies for Healthcare (pp. 477-495). Cham: Springer Nature Switzerland.

[10] Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C. and McTear, M., 2017, July. Towards a chatbot for digital counselling. In Proceedings of the 31st international BCS human computer interaction conference (HCI 2017). BCS Learning & Development.

[11] Spitzer, R.L., Kroenke, K., Williams, J.B. and Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine, 166(10), pp.1092-1097

[12] Kroenke, K., Spitzer, R.L. and Williams, J.B., 2001. The phq-9. Journal of general internal medicine, 16(9), pp.606-613.

[13] Cohen, S., Kamarck, T. and Mermelstein, R., 1994. Perceived stress scale. Measuring stress: A guide for health and social scientists

[14] Saunders, J.B., Aasland, O.G., Babor, T.F., De la Fuente, J.R. and Grant, M., 1993. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. Addiction, 88(6), pp.791-804.

[15] Espie, C.A., Kyle, S.D., Hames, P., Gardani, M., Fleming, L. and Cape, J., 2014. The Sleep Condition Indicator: a clinical screening tool to evaluate insomnia disorder. BMJ open, 4(3), p.e004183

[16] Agrawal, R., and Srikant, R., 1994. Fast Algorithms for Mining Association Rules. Proceedings of the 20th International Conference on Very Large Databases (VLDB), 487-499.

[17] Nahum-Shani, I., Smith, S.N., Spring, B.J., Collins, L.M., Witkiewitz, K., Tewari, A. and Murphy, S.A., 2018. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Annals of Behavioral Medicine, pp.1-17.

[18] Versluis, A., Verkuil, B., Spinhoven, P., van der Ploeg, M.M. and Brosschot, J.F., 2016. Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. *Journal of medical Internet research*, *18*(6), p.e5642.