Recommendation Modeling for Health Self-Management Applications for People with Rheumatoid Arthritis

Nadine Schwab[1*,2], Günther Zauner[1], Christoph Urach[1],
Paul Studenic[3,4], Helga Radner[3], Nasim Nakhost-Lotfi[3], Tanja Stamm[5],
Thomas Hammer-Jakobsen[6], Andreas Dam[6,7], Niki Popper[1,2,8]

[1]dwh simulation services & technical solutions, dwh GmbH, Neustiftgasse 57-59, 1070 Vienna, Austria
*nadine.schwab@dwh.at
[2]Institute for Information Systems Engineering, TU Wien, Favoritenstraße 9-11, 1040, Vienna, Austria
[3]Division of Rheumatology, Department of Internal Medicine III Medical University of Vienna, Vienna, Austria
[4]Division of Rheumatology, Department of Medicine (Solna), Karolinska Institutet, Stockholm, SWEDEN
[5]Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria
[6]HealthBuddy, Vesterbrogade 149, 1620 Copenhagen, Denmark
[7]Daman, Vesterbrogade 149, 1620 Copenhagen, Denmark
[8]Inst.of Statistics and Math. Methods in Economics, TU Wien, Wiedner Hauptstraße 8, 1040, Vienna, Austria

SNE 36(2), 2026, 99-106, DOI: 10.11128/sne.36.tn.10776
Selected ASIM SST 2024 Postconf. Publ.: 2025-06-10
Received Revised: 2025-12-15; Accepted: 2025-12-20
SNE - Simulation Notes Europe, ARGESIM Publisher Vienna
ISSN Print 2305-9974, Online 2306-0271, www.sne-journal.org

Abstract. For the containment of chronic diseases, mHealth tools, for example mobile apps, provide great opportunities to track the disease progression and to give useful recommendations. In this project, possible data-driven enhancements for an already existing mobile app for patients with Rheumatoid Arthritis are discussed, developed and implemented. This happens in an ongoing feedback-cycle, including app developers, medical experts, patients and data scientists. The new features improve the app experience and are currently being evaluated in an observational study.

Introduction

Rheumatoid Arthritis (RA) is a chronic autoimmune disease, which affects over 3 Million people in Europe. People dealing with RA usually suffer from painful and tender joints. This and the whole plethora of also more subtle symptoms like fatigue strongly impacts their every day life [1], [2]. There is no cure for RA, but the symptoms and the progress of the disease can nowadays be halted effectively and improved by targeted pharmacological and non-pharmacological interventions. mHealth tools provide opportunities for the tracking and self-management of symptoms on a patient-individual basis [3].

RheumaBuddy is such an app that serves as an electronic diary for RA-specific symptoms: Users can track their symptoms, such as Mood, Pain, Fatigue and Stiffness, as well as activities such as walking or sleeping, on a daily basis. This helps keeping track of their disease progression, in general or for the next doctor appointment.

Throughout the project RheumaBuddy4.0 (RB4.0), it is the aim to improve this app. In order to achieve this, firstly, an extensive data analysis has been performed to figure out which additional features are possible and make sense in the current app. Next, these data-driven recommendations are implemented in the app and tested by the users. The whole process takes place in a permanent feedback loop including the Danish app developers DAMAN, medical experts from the Medical University of Vienna (MUV), data scientists from dwh GmbH, and RA patients who use and test the app. Fig.1 shows a schematic depiction of the workflow in the project.

1 App Status at Begin of Project

At the beginning of the project, RheumaBuddy is already available as a mobile app in multiple language with German, English, and Danish being the most popular ones. Users have the possibility to enter RA-related symptoms, such as Mood, Pain, Fatigue and Stiffness as well as self-reported symptoms, where they could track whatever they are interested in.

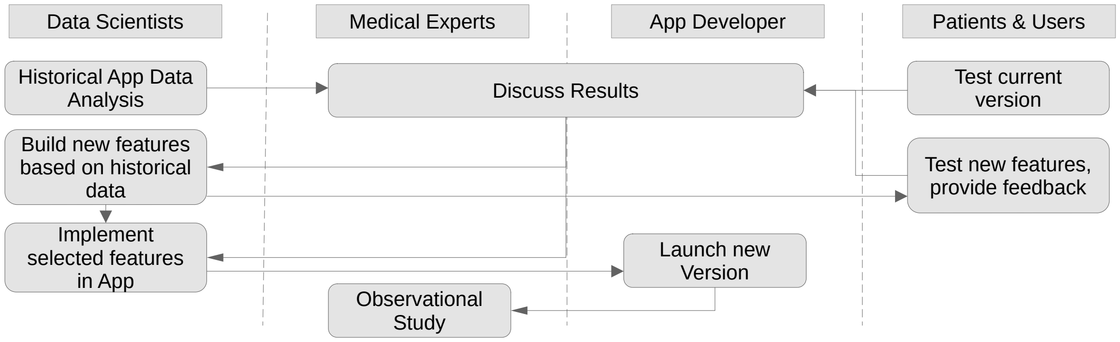


Figure 1: A schematic depiction of the general workflow of the project. The development of new features and the discussion with the experts is often repeated multiple times.

Every symptom can be scored on a smiley scale indicating the extent of the respective symp-tomatology. These smileys rank from very happy to very unhappy in five steps. Users can also enter acti-vities, such as walking or sleeping, and how long these activities took on that day. Furthermore, they can document their joint pains on a body map. And finally, they can write free-text in a diary, as well as use the community to ask or to answer to questions.

The community is also available in different languages. The user's status and progression can always be seen on the Status page, where line graphs visualize the development of all the values tracked by the user.

The goal of the project is to implement a virtual coach within the app. This coach should provide users entering sufficient data into the app with personal recommendations based on their individual entries. This goal comes with the following challenges:

- **Amount of regular data:** In order to provide useful, data-based information, a sufficient amount of data per user is required, i.e. the users must enter data on a regular (ideally daily) basis over a few weeks, at the very least. Usually, it is difficult to motivate app users to provide this amount of input.
- **Subjectivity of app entries:** The app entries are subjective for every user, and it is not trivial to compare them. E.g., for one user, "good mood" means they actually had an amazing day, while for another user, it only means that everything is alright.
- **Not measurable disease parameters:** There are hardly any RA-specific parameters that can be measured objectively on a remote basis, i.e. without a medical expert. As opposed to e.g. Diabetes, where blood sugar can be objectively measured and provides a good indicator for the current disease activity, rheumatologists can only evaluate a patient's disease status by assessing the swollen, hurting and tender joints of a patient besides deviations in inflammatory parameters, blood levels, or changes in bone and joint structure with imaging modalities.

2 Data Analysis

An extensive data analysis was conducted in order to determine which app enhancements are possible with the available data, and how the results look like. The results were frequently discussed with the app developers and medical experts. Furthermore, feedback was constantly obtained from app users. Based on these feedback loops, possible app features are discussed and designed.

The data analysis with its results is presented in this chapter. The next chapter describes a new app feature, which evolved from this analysis.

2.1 User Behavior

First we analyzed, how extensively the app is used by the users. We analyzed the development of the number of unique users per day (which means they entered at least one value into the app on that day) and which entry types are the most popular.

The development of the users can be seen in Figure 2. A peak is visible in October 2020 as a result of a big marketing campaign. The number of users didn't stay that high, but remained significantly higher than before.

The most popular entry types can be seen in Figure 3. The most popular ones are the four main symptoms Mood, Pain, Fatigue and Stiffness, followed by the Pain Map. The diary and activity tracking is far less popular, as well as the self-reported symptoms. The latter shows that the four main symptoms seem to be sufficient for most app users. The least popular is writing and reading in the community.

Possible reasons for this heterogeneous behaviour were discussed with the consortium, especially with some app users. The users claimed that for those entries that are used the most often, it is much easier and faster to enter values than for the others. It goes without saying that writing into the diary or community is more time-consuming than choosing a smiley value for a symptom. The activity values, on the other hand, are not trivial to measure. Users claimed that it is hard to remember in the middle of the day, how long exactly they slept. Considering "Motion", it was not clear to them, what accounts to this, e.g. whether a walk counts already, or only exhausting physical activities, which is why they soon stopped entering values there.

2.2 Numeric Correlations

We investigated the correlations between the entries of the users. The entries are obtained directly from the app: As mentioned in section 1, a user can enter their symptoms via a smiley, which ranks from happy to unhappy in five steps. These steps are denoted in the back-end with the integer values 1 to 5 (1 being the most unhappy one), which are used to compute correlations. Considering activity values (Sleep, Motion, Working hours), the actual time spans in seconds are used. These correlations are displayed in Figure 4. Strong corre-lations can be seen between the four main symptoms, but not between the rest of the entry types. However, no correlations between symptoms and activities might also be because the users had problems entering reasonable values for these fields, as discussed in Subsection 2.1.

2.3 Free-Text Analysis

Although these types of entries are less popular, we also analyzed the free-text entries, i.e. diary, community, and self-reported symptoms. We wanted to grasp what the users are writing about (besides the pre-defined entries), and whether some information that is relevant for all users can be extracted from the entries. To achieve this, we firstly translated all self-reported symptoms, diary and community entries to English language (using the Python Library Deep Translator), removed fill- and stop-words (using [7]) and counted which words appeared the most often in the entries. This is visualized in Figure 5.

As it can be seen immediately in the Subfigures 5b and 5a, the most important topics in the diary and in the community are RA-related symptoms or medication. Considering the self-reported symptoms in Subfigure 5c, interestingly, four of the five most frequent ones are exactly the four main symptoms that are provided in the app anyway, i.e. Mood, Pain, Fatigue and Stiffness. We assume, the users entered their data in the self-reported symptoms due to a misunderstanding - however, this indicates again that these symptoms seem to be of utmost importance. However, as the remaining symptoms (resp. the self-reported symptoms in general) are used by very few users, no further analysis was conducted in this area.

2.4 Community Crawling

The community is open for every RheumaBuddy-user, but as it was shown in Figure 3, there are only few users that use the community - even just for reading. Figure 5b indicates that many community posts and comments deal with RA-related symptoms and medication. As the community is built in a question-and-answer-style, i.e. a user can ask a question and any other user can answer to them, we assume that there are many recommendations hidden in the community. These recommendations would be of great value for many RheumaBuddy users, not only those who read and write in the community.

Therefore, we wanted to systematically collect helpful recommendations from the community. Depending on the information that could be gained, these could be forwarded to all users, or only to specific ones where it might fit (depending on the entries the users make). We collected recommendations using the following, semi-automated approach:

1. Filter the community posts by posts which ask for advice on something, using regular expressions.
2. Retrieve all answers to this post and remove all answers from the user who asked the question.
3. Read the answers in combination with the question and decide whether to keep or to discard it.

A fully-automated approach was not developed within the scope of this project, as the Community Crawling was only experimental, and the amount of data available in the community is too small to justify much research in the field of automated text mining. Nevertheless, this semi-automated approach gravely reduced the manual work and enabled to quickly obtain a document with possible recommendations collected from the community.

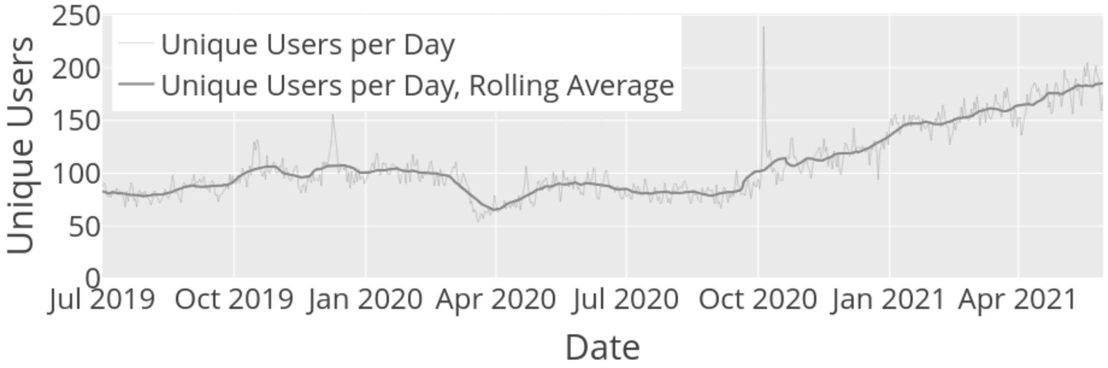


Figure 2: Plot of the development of the number of users per day. The thick line depicts a seven-day rolling average, the thin line the actual values. A peak can be seen as the result of a big marketing campaign, which led to a general increase of users per day.

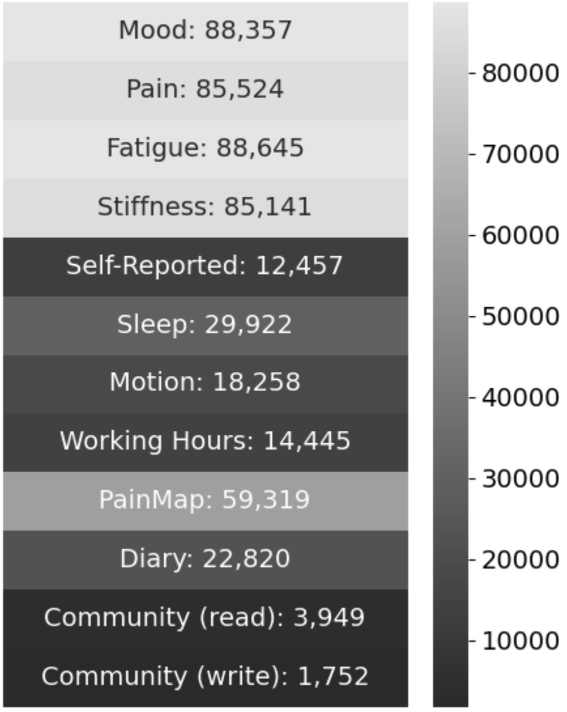
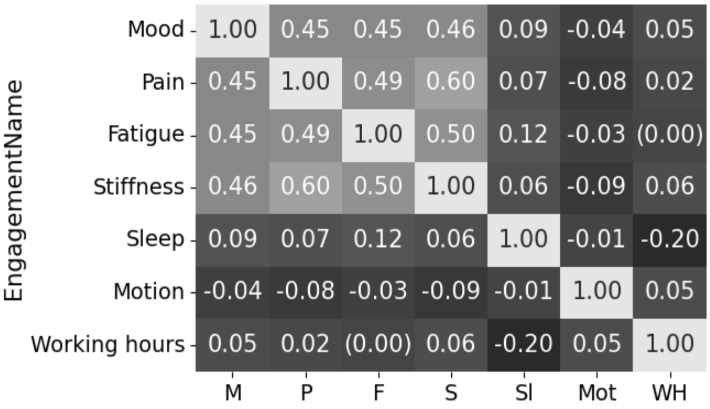
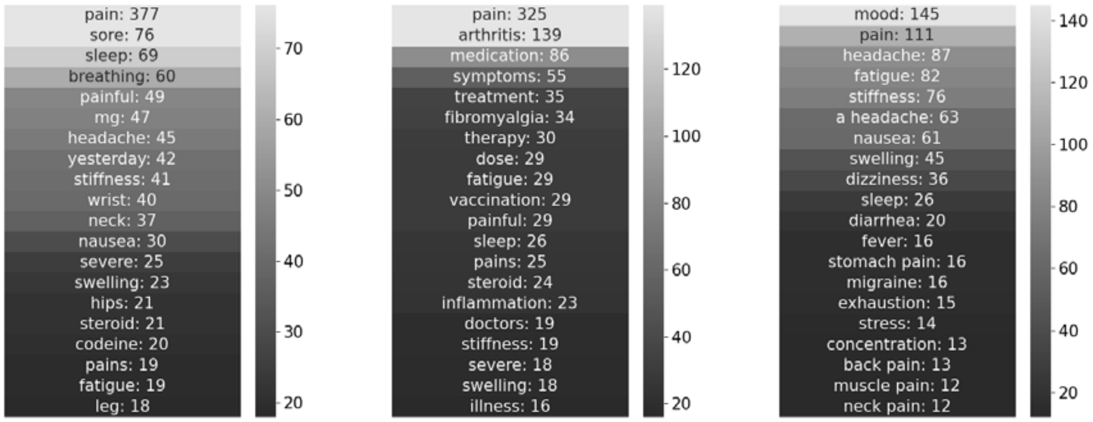


Figure 3: Depiction of the total count of entries for each entry type. The fast-to-enter symptoms as well as the pain map are the most popular entry types.



	M	P	F	S	Sl	Mot	WH
Mood	1.00	0.45	0.45	0.46	0.09	-0.04	0.05
Pain	0.45	1.00	0.49	0.60	0.07	-0.08	0.02
Fatigue	0.45	0.49	1.00	0.50	0.12	-0.03	(0.00)
Stiffness	0.46	0.60	0.50	1.00	0.06	-0.09	0.06
Sleep	0.09	0.07	0.12	0.06	1.00	-0.01	-0.20
Motion	-0.04	-0.08	-0.03	-0.09	-0.01	1.00	0.05
Working hours	0.05	0.02	(0.00)	0.06	-0.20	0.05	1.00

Figure 4: Linear, numeric correlations between the entry types based on 456,998 observations. Correlations between the four main symptoms are present, symptoms and activities seem to be uncorrelated.



(a) Most frequent words in the diary. (b) Most frequent words in the community. (c) Most frequent Self-Reported Symptoms.

Figure 5: Heatmaps for the most often used words in all kinds of free texts. The main topics for all sources are disease-related, i.e. symptoms and medication are the most popular.

However, the obtained recommendations cannot be used as they are as app recommendations. These recommendations are mainly tips from the users who usually do not have any medical background. Sharing what might help one user does not necessarily help another user, but could in the worst case actually harm them. And while it is fair to have recommendations in the community that might not be suitable for all users, app-provided recommendations must in any case be medically solid and do no harm to the users. Therefore, the recommendations obtained via community crawling cannot automatically be used in the app.

Still, the recommendations can be collected and further used. For example, they can be shared with medical experts which can assess on a patient-individual basis whether some of the recommendations could be helpful for some of their patients. Or specific medical studies could be started, proving (or disproving) some of the interesting recommendations. And finally, the developed method can be used in different contexts, where obtaining unsuitable recommendations is not that problematic.

2.5 User Clustering

Another idea was grouping similar users, based on their entries. Having such a group of users, the app could e.g. provide specific recommendations, based on the group characteristics. Users in such a group could also be brought together in an anonymous chat (if they consent) and share their experiences.

To achieve this, we measured the similarity between the main symptom entries (Mood, Pain, Fatigue and Stiffness) of the users. In order to recognize similar patterns even if they are shifted in time or slightly distorted, we measured the similarity using Dynamic Time Warping (DTW, [5]). We focused on one of the symptoms at a time and computed for all users with at least 10 entries the pairwise similarities between their values using DTW. Next, we grouped the users based on their pairwise distances. We tried different clustering algorithms using Scikit-Learn [6], finding that Spectral Clustering provided the most promising algorithm.

However, on the available app data, it was not possible to find useful clusters. Hardly any user provides daily entries, but rather one entry every few days. Furthermore, the values are discrete values between 1 and 5, which resulted in entry curves where no sensible pattern could be found. With the available values, it is not possible to recognize users that get steadily better or worse over some time. To achieve this, a scale where more precise entries are possible would be required.

3 User-individual Correlations

We figured that the users of the app might not be interested in the development of their values, but also in the relationship between these values. As we showed in Subsection 2.2, significant correlations between the symptoms are present when considering all users at once.

To our surprise, no correlations between symptoms and activities could be found. However, during our consortium discussions, also including patient research partner, we agreed that relationships between activity patterns and symptoms should exist. Not finding any with the current approach could be due to several reasons:

- As discussed in Subsection 2.1, the users had problems entering reasonable values to the activity functions in the app, and as can be seen in Figure 3, the number of entries for activity values is much smaller.
- Users with positive and negative relationships could even each other out when being considered all at once.
- Also when it comes to one user, they don't necessarily have a *linear* relationship between their activities and symptoms, but rather some level of activity where they feel best/worst.

We decided to have a closer look at user-individual relationships between the values, and to not only take into account linear correlations. Additionally, the relationships between a user's individual input values is probably far more interesting to them than the correlations of all users. Therefore, we investigated on a user-individual basis, between which input values we could find the strongest significant relationships.

3.1 New Activity Values

First, we wanted to overcome the problem with entering reasonable activity values into the app. To achieve this, a step counter was added to the application. Of course, this does not cover the whole variety of physical activity (a step counter does not distinguish between walking and jogging, for example), but these values are added to the app automatically and provide a good approximation of the user's level of physical activity. Furthermore, the smiley values were enhanced by activity categories, enabling a user to provide input on a scale of 1 to 5, how physically active they were that day.

3.2 Model

We focus on the relationships between symptoms and activities per user, and consider three different kinds of relationships that could occur between the values:

- Linear relationships, e.g. "The more you walk, the better is your mood."
- Cut-Off values, e.g. "Your mood increases signifi-cantly when you walk around 6,000 steps per day."
- Extreme values, e.g. "Your mood is best when you walk around 6,000 steps per day."

These relationships are computed twice for each symptom-activity-pair, between same-day data as well as between activity data on one day and symptom data the next day. Thus, a possible output could also be "The more you walked *the day before*, the better is your mood." Additionally, every relationship can also be negative, e.g. "The *less* you walk, the better is your mood."

To achieve this, we firstly retrieved a user's recent data from their last 30 entry days. If they provided values for the considered symptom-activity-combination on at least on 10 different days (i.e. if we have at least 10 entry pairs), we fitted the user-data with three different models: A linear regression model (LR), a decision tree (DT), and a piece-wise linear function (PWLF).

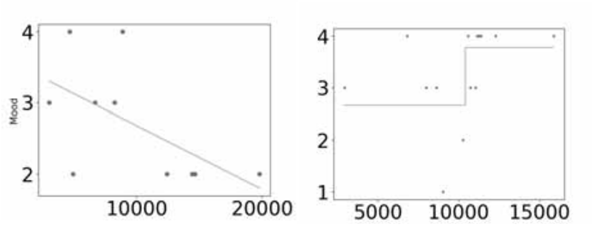
Linear Regression Model. We fitted the data with a linear regression curve of the form

$$y_i = a \cdot x_i + b + r_i$$

with y_i being the symptom values, x_i the activity values and r_i the residuals. a is the slope of the function, which is later used to determine if there is a positive or negative relationship between the two values.

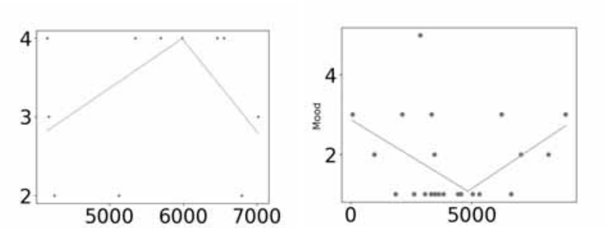
Decision Tree. We also fitted the data with a decision tree with one split node and constant values at the leaves.

Piece-wise linear function. If the user provided at least 15 value pairs in the given time frame, we also fitted the data with a continuous piece-wise linear function with one split, using the python library PWLF [8]. In order for this model to be valid in our context, the slopes of the two linear functions must have different signs, i.e. the break point must be a minimum or a maximum value. Additionally, the break point was only allowed within the inner 60% of the data range, and both linear functions must be supported by at least 5 data points. Otherwise, the PWLF would always fit the data better than the other two models without providing additional information.



(a) A user with a negative linear correlation between Mood and Steps.

(b) A user with a cut-off value in the relationship between Mood and Steps.



(c) A user with a maximum value in the relationship between Mood and Steps.

(d) A user with a minimum value in the relationship between Mood and Steps.

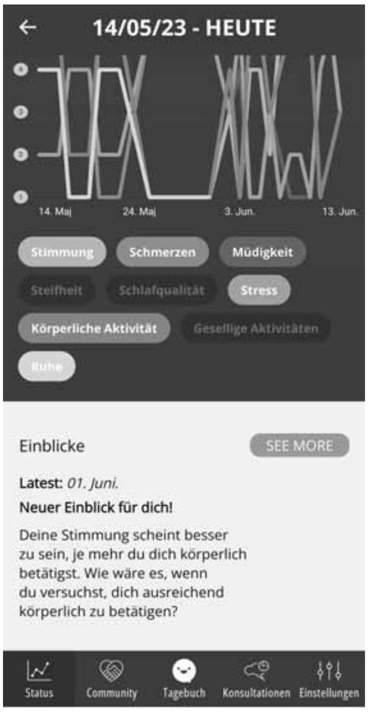
Figure 6: Example plots of four users, depicting their individual relationships between number of steps and mood. The rather simple models still provide an interpretable relationship between the values.

Figure 6 shows some sample plots of the actual values of four different users, together with the model that fits their data. Note that for Subfigures 6c and 6d, only a very weak linear relationship would have been found. From each of these models, the R^2-Score is computed as a measure of fitting, and the model with the highest R^2 is kept. Thus, for every symptom-activity-combination, we compute one relationship. Finally, from all the symptom-activity-combinations, we choose the one with the highest R^2-Score and display it as an insight to the user.

3.3 Deployment

This feature was added into the RheumaBuddy app. It was deployed via a Docker Container [9] which runs on the server of the app provider DAMAN. Thus, the described relationships are computed in real-time and are communicated live to the users. A sample output from the app can be seen in Figure 7.

However, the number of entries required to obtain a significant relationship is quite high, and only few users entered enough values to obtain such a recommendation.



Einblicke

Latest: *01. Juni.*

Neuer Einblick für dich!

Deine Stimmung scheint besser zu sein, je mehr du dich körperlich betätigst. Wie wäre es, wenn du versuchst, dich ausreichend körperlich zu betätigen?

Figure 7: An actual recommendation (in German language) as provided from the **RheumaBuddy** app, stating that this user's mood seems to be better the more physically active they are.

4 Validation Process

The new app features are validated concerning face validity from the app developers. They check whether the app feedback makes sense and fits to the provided user input. Furthermore, they make sure that the insight is prompted to the users, and to how many. A selection of patients from their company review board tests the new features and provides thorough feedback on the functionality and the insights.

However, validation is at this point not complete and refinements would potentially be tackled in several iteration rounds. For putting **RheumaBuddy4.0** on the track for CE marketing, certain aspects of a final software product would need to be submitted.

The validation process can be optimized by integrating results from the observational study at the Medical University of Vienna ("Successful Implementation of an RA management app - the IRAMA study" with ethical approval number: EK-Nr: 1846/2022) which is in process. The major aims of the IRAMA study concern the assessment of the RheumaBuddy mobile app quality, using a standardized questionnaire to evaluate patients' experiences, preferences and needs for the use of the **RheumaBuddy** App in semi-structured telephone interviews, and finally to gain more detailed insight on patients' experiences, needs and preferences in semi-structured focus group sessions.

The data of this study is directly linked with a clinical registry and the app information the users are collecting. Thus, the foundation for a thorough validation via an observational study is already in place.

5 Discussion and Outlook

Overall, it can be stated that the project is progressing successfully. An interesting, helpful feature for the app users has been implemented, and we gained a lot of insights by means of the data analysis and the discussions in the consortium. We expect further interesting insights when the observational study is finalized.

One of the main challenges of the project was the lack of user input into the app. While the functions work well and can give useful insights to the users, hardly any users obtain such recommendations, because they stop entering values into the app too early. Some ways of Gamification to increase the motivation of the app users could be considered, e.g. counting how many days in a row they entered something into the app, and motivating not to break this "streak" - this has been shown to be helpful in many mobile apps [10].

The combination of passively tracked data with traditional or short PROMs as used in RheumaBuddy is potentially more promising to be supportive regarding improvement of the quality of life of people with RA when combined with the treating rheumatology team. As RA can affect people of any age, the background, i.e. experience with technology or eHealth literacy, is very heterogeneous, which makes it challenging to develop digitally assisted technologies that support every patient on an equal basis. Technology has been utterly fast paced over the past years, sharing breaking record news multiple times per year, but medical systems and people work and live on a slower pace. Thus, we would need to harmonise the slopes of development with the needs and opportunities that appear for implementation of novel technologies. This will bring benefit to people living with chronic conditions like RA as well as health care providers and system payers [4].

Acknowledgement

The project RB4.0 (FFG project number 880456) was funded by the Austrian Research Promotion Agency (FFG), the national funding agency for industrial research and development in Austria, as part of the call for proposals *Eurostars-2 CoD 13* (www.ffg.at/en).

Publication Remark

This contribution is the revised version of the conference version published in ***ASIM SST 2024** Munich Tagungsband Langbeiträge,* ARGESIM Report 47, e-ISBN978-3-903347-65-6,Vol. DOI 10.11128/arep.47, Article DOI: 10.11128/arep.47.a4725, p 43-50.

References

[1] Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, Scott D, Silman A. *The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century.* Rheumatology (Oxford). 2002 Jul; 41(7):793-800.

[2] Finckh A, Gilbert B, Hodkinson B, Bae S-C, Thomas R, Deane KD, Alpizar-Rodriguez D, Lauper K. *Global epidemiology of rheumatoid arthritis.* Nature Reviews Rheumatology. 2022 2022/09/06 doi:10.1038/s41584-022-00827-y.

[3] Winthrop KL, Mease P, Kerschbaumer A, Voll RE, Breedveld FC, Smolen JS, Gottenberg J-E, Baraliakos X, Kiener HP, Aletaha D, Isaacs JD, Buch MH, Crow MK, Kay J, Crofford L, Vollenhoven RFv, Ospelt C, Siebert S, Kloppenburg M, McInnes IB, Huizinga TW, Gravallese EM. *Unmet need in rheumatology: reports from the Advances in Targeted Therapies meeting, 2023.* Annals of the Rheumatic Diseases. 2023:ard-2023-224916 doi:10.1136/ard-2023-224916.

[4] De Cock D, Myasoedova E, Aletaha D, Studenic P. *Big data analyses and individual health profiling in the arena of rheumatic and musculoskeletal diseases (RMDs).* Ther Adv Musculoskelet Dis. 2022; 14:1759720X221105978 doi:10.1177/1759720X221105978

[5] Springer Berlin Heidelberg. Dynamic Time Warping. In: *Dynamic Time Warping*. 2007. p 69–84.

[6] *Scikit-learn: Machine Learning in Python*, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[7] Bird S., Klein E., Loper E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc."

[8] Jekel Charles, Venter Gerhard. (2019). *pwlf: A Python Library for Fitting 1D Continuous Piecewise Linear Functions.* 10.13140/RG.2.2.28530.56007.

[9] Cito Jürgen, Ferme Vincenzo, Gall Harald. (2016). *Using Docker Containers to Improve Reproducibility in Software and Web Engineering Research.* 609-612. 10.1007/978-3-319-38791-8_58.

[10] Galetta, Giuseppe. (2013). *The Gamification: Applications and Developments for Creativity and Education.* 10.13140/RG.2.2.24817.68965.