Grant Agreement No.: 823783
Call: H2020-FETPROACT-2018-2020
Topic: H2020-FETPROACT-2018-01
Type of action: RIA

WENET
INTERNET OF US

D4.1 ALGORITHMS AND IMPLEMENTATION OF SINGLE USER INCENTIVE DESIGN

Revision: v.1.0

Work package	WP 4
Task	Task Number T4.1
Due date	30/04/2020
Submission date	28/04/2020
Deliverable lead	Avi Segal
Version	1.0
Authors	Kobi Gal, Avi Segal (BGU)
Reviewers	Giuseppe A. Veltri (UNITEN)

Abstract	We describe our work of developing non-monetary incentives for increasing motivation and improving productivity and engagement for single users. We elaborate on our research and implementations of harnessing a recommendation based method focusing on increasing engagement and satisfaction by matching tasks to users. In our research we take a computational based approach, which uses participants' interaction histories in the system to predict their future responses to considered incentives. Our implementations and trials point to the potential of these approaches as well as to the need to better tailor them to a diversified population, a task we will undertake in our next steps. During this research we worked with systems in a large scale volunteer-based crowdsourcing domain (citizen science).
Keywords	Incentive Design, Recommendation Systems

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.8	08/04/2020	Version sent to WP4 internal review	Avi Segal (BGU)
v0.9	12/04/2020	Version sent to WeNet internal review	Avi Segal (BGU)
V1.0	28/04/2020	Version ready for external review	Avi Segal (BGU)

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable:	**R**	
Dissemination Level		
PU	Public, fully open, e.g. web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to WeNet project and Commission Services	

** R: Document, report (excluding the periodic and final reports)*

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

EXECUTIVE SUMMARY

We report on our work of developing non-monetary incentives for increasing motivation and improving productivity and engagement for single users in online systems. We elaborate on our research and implementations of a recommendation based method focusing on increasing engagement and satisfaction by matching tasks to users. In our research we take a computational based approach, which uses participants' interaction histories in the system to predict their future responses to considered incentives.

Our recommendation based work matches thousands of volunteers with Citizen Science projects. Citizen science engages people in scientific research. As technology develops, the scale of citizen science projects and members dramatically increases. For example, SciStarter, our partner in the recommendation based project, offers more than 3,000 projects and recruits volunteers through media and other organizations, bringing citizen science to people. Given the sheer size of available projects, finding the right project, which best suits the user preferences and capabilities, has become a major challenge and is essential for keeping volunteers motivated and active.

We develop a novel approach for supplying the right recommendation to volunteers based on their profile and logged activities. We use methods from the field of recommendation systems, including collaborative filtering methods and matrix factorization. This research develops a new approach for personalized recommendations in the SciStarter ecosystem, in order to increase their engagement and motivation to contribute to citizen science platforms. This work lays the foundation to enhance participants' motivation and learning in online systems by aligning interests with projects through AI recommendation tools.

This recommendation system was deployed in the SciStarter ecosystem, and was evaluated in an online study involving thousands of users who were informed about participating in a study involving AI based recommendation of new projects. Volunteers were randomly divided into different cohorts, which varied the recommendation algorithm that was used to generate suggested projects. We were able to show that our recommendation system was able to 1) engage people in new projects that they had never tried before; 2) led to increased participation in SciStarter projects; 3) the cohort of volunteers receiving recommendations created by the SVD algorithm (matri factorization) exhibited the highest levels of contributions to new projects, when compared to the other cohorts. This is the first study using AI based recommendation tools in large scale citizen science platforms.

We analyse our results in connection to multiple diversity aspects of our target population and notice differences in the impact of our algorithms on different diverse groups. This leads to a key goal of our next steps – considering diversity features explicitly as part of our machine learning algorithms. Additionally, our next steps will seek to combine badges and recommendations and develop a multi-incentive system for integration within the WeNet platform.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ML	Machine Learning
AI	Artificial Intelligence
RS	Recommendation System

Co-funded by the Horizon 2020
Framework Programme of the European Union

1 INTRODUCTION

Citizen science engages people in contributing to the science by collecting, categorizing, transcribing, or analyzing scientific data [4, 11, 5]. These platforms offer thousands of different projects which advance scientific knowledge all around the world. With this wide variety of projects, finding the right project, which best suits the user preferences and capabilities, has become a major challenge [22, 6] and is essential for keeping the user motivated, active and satisfied. In this research we aimed to test how a personalized AI-based recommendation system can address this challenge and incentivise users, increase their motivation and satisfaction with the projects they choose to contribute to.

Recommendation systems have been used in other domains, such as e-commerce, news, social media, content aggregation and censorship systems [14, 8]. Personalized recommendations have been shown to constitute strong incentivising mechanisms in monetary based systems, were they are used and studied extensively [29]. Specifically, past work has shown that personalized recommendations are incentivising and leading to increased motivation, higher engagement levels, extended consumption level and increased purchasing behaviour [31,32]. Additionally, applying personalized recommendations in online educational settings has shown promising impact on the motivation and learning gains of target learners [30]. Thus, such mechanisms are deemed important in the context of a diversity focused social project such as WeNet, were non-monetary policies are to be used for incentivising users and achieving system goals.

This research develops a new approach for personalized recommendations in citizen science platform, in order to increase their engagement and motivation to contribute. Our recommendation system delivers personalized recommendations to signed-in users by recommending them with new projects based on their past history on the site and based on projects' characteristics. We apply this approach on data that is obtained from the SciStarter platform (https://scistarter.org). SciStarter offers more than 3,000 projects and recruits volunteers through media and other organizations, bringing citizen science to people. Our work emphasises the value of intelligent recommendations in large-scale online systems. In such domains, the majority of users are characterized with very low engagement [2], where both the duration and the quantity of their contribution activities decreases significantly over time. Our results points to the potential of the proposed recommendation based approach and algorithms to incentivize users in voluntary non-monetary domains, such as the WeNet system. Our future work will extend these algorithms to better cope with a diversified user base as well as add other incentivizing mechanisms in addition to personalized recommendations.

2 PAST WORK

This work relates to past work in citizen science research as well as recommendation systems. First, we will discuss work from the field of citizen science, and the use of AI in this domain in order to increase users engagement. In the second part of the review, we will describe in detail the field of recommendation systems and their applications.

2.1 CITIZEN SCIENCE – MOTIVATION AND LEVEL OF ENGAGEMENT

Online participation in citizen science projects has become very common [18]. Yet, most of the contributions rely on a very small proportion of participants. This group of participants is highly engaged with the projects and make thousands of tasks. However, in most citizen science projects, the majority of participants carry out only a few tasks. Many researches have explored the incentives and motivations of participants in order to increase participants engagement. Kragh et al. [15] claimed that participants in citizen science projects are motivated by personal interest and desire to learn something new, as well as by the desire to volunteer and wish to contribute to science. Raddic et al. [20] support the above claim and discuss that participants engagement is mostly originated in pure interest in the project topic, such as astronomy. Based on this finding, we also implemented a hybrid algorithm which combines data of projects topics. Arazy et al. [18] have explored this by separating the question to the quantity of contribution and quality of contributions. They showed that quantity of contribution is mostly determined by the user interest in the project and by social norms. In contrast, the quality of contribution is determined by understanding the importance of the task and by the user's reputation. In our work, we aim to increase both the quantity and quality of contributions, based on the described factors.

Significant prior work was done in order to increase participants engagement, which is influenced by the discussed motives. Segal et al. [27] have developed an intelligent approach which combines model-based reinforcement learning with off-line policy evaluation in order to generate intervention policies which significantly increase users' contributions. Laut et al. [17] have demonstrated how participants are affected by virtual peers and showed that participants' contribution could be enhanced through the presence of virtual peers.

Ponciano et al. [19] characterized volunteers' task execution patterns across projects and showed that volunteers tend to explore multiple projects in citizen science platforms, but they perform tasks regularly in just a few of them. They have also shown that volunteers recruited from other projects on the platform tend to get more engaged than those recruited outside the platform. This finding is a great incentive to increase user engagement through platform instead of targeted projects like we do in our research.

In this research, we attempt to enhance participant engagement to citizen science projects by recommending the user projects which best suit the user preferences and capabilities

2.2 RECOMMENDATION SYSTEMS

Recommendation systems aim to provide the most relevant items to users by learning their choices and producing the results that co-relates to their interests and needs [21].

Co-funded by the Horizon 2020 Framework Programme of the European Union

Generally, recommendation systems are based on two main approaches: a content-based approach and collaborative filtering approach.

The content-based strategy [3] is based on shared items and users characteristics. For example, project characteristics can include the project task, location and targeted age group. The user is recommended with items according to these characteristics. Alternatively, collaborative filtering (CF) strategy [25] is based only on user past activities, and recommendation is based on analyzed relations between users and items. Recommendations based on CF rely on different types of input. Preferably, explicit data, which is based on some sort of user-item rating. Generally, this type of input is not available in many domains. Therefore, implicit data is taken into consideration, where user-item interactions are inferred by user behaviour, such as participation, clicks, search patterns etc. A good discussion on collaborative filtering based implicit feedback is given by Hu. et al. [13]. Hu. et al. discuss the unique characteristics of implicit data, which justify the replacement of explicit based algorithms. In the educational domains, other works also tried to increase users' engagement and participation by personalized recommendations. Labarthe et al. [7] built a recommender system that recommends relevant and rich-potential contacts based on user profile and activities. They have shown that by recommending this list of contacts, students were much more likely to persist and engage in MOOCs. A subsequent work of Dwivedi et al. [16] has developed a recommender system that recommends online courses to students based on their grades in other subjects. This recommender was based on collaborative filtering techniques and particularly item based recommendations.

Some other works that concern user engagement with recommendation systems have shown how early intervention significantly increase user engagement [9]. Wu et al. [28], have shown how tracking user's clicks and return behaviour succeeds to increase user engagement with the recommendation system. They have formulated the optimization of long-term user engagement as a sequential decision making problem, where a recommendation is based on both the estimated immediate user click and the expected clicks results from the users' future return.

All the systems described above are designed for different goals. In this work, we introduce a recommendation system with the goal of increasing user engagement and satisfaction in citizen science projects. This system will take into consideration the special factors relevant to this domain, and by this, we develop a new strategy for keeping the user engaged with citizen science projects.

Co-funded by the Horizon 2020
Framework Programme of the European Union

3 EXPERIMENT DESIGN

The research question we study is whether intelligent recommendation of citizen science projects will result in increased engagement and improved contributions as compared to existing SciStarter tools. In order to address our research question, we compare several recommendation algorithms that are based on existing state-of-the-art tools, adapted to the citizen science domain.

3.1 INPUT DATA

During the generation of personalized projects recommendations we consider a number of data sources. (1) Interactions with affiliate projects: data generated as a result of users' activities with affiliate projects, e.g. joining a project, making a contribution to a project or participating in a project. An affiliate project is one that uses a specific API to report back to SciStarter each time a logged in SciStarter user has contributed data or analyzed data on that project's website or app. (2) Direct interactions with projects on Scistarter's website; any type of click a user performs in Scistarter is recorded. The clicks that are considered as an interaction with a project are any clicks related to the project - could be searching the project, filling a form about the project, etc. The data is represented as a binary matrix $R[U * I]$, where U is the number of users, and I is the number of projects. $R[i, j] = 1$ indicates that user i had some interaction with project j, and $R[i, j] = 0$ otherwise. (Interaction is either an interaction with affiliate project or interaction with projects on SciStarter). Consequently, our system is a top-N recommender system rather than a predictor.

3.2 ALGORITHMS

We design four different algorithms: CF user based, CF Item based, Matrix Factorization - SVD, Popularity based.

3.2.1 User Based Collaborative Filtering

In this algorithm, the recommendation is based on user similarities [25]. The ranking of a project for a target user is computed by comparing users who have done similar projects. For example, the algorithm recommends the iNaturalist project (an online biodiversity project offered in SciStarter) to user A, because similar users to the target user have contributed to this project.

We use KNN algorithm to find similar users. We choose K = 100, which we found to be the optimal k for this problem. At first, we encounter a problem by choosing this K, since in our data, we have users who act exactly the same. Therefore, when we try to find the most similar projects of k similar users to user i, we get exactly the projects user i has already participated in. For this reason, we increase the value of K to K + 100 in each iteration, until we have a sufficient variety of projects from similar users (where similar is now a more broader definition).

3.2.2 Item Based Collaborative Filtering

In this algorithm, the recommendation is based on projects similarities [25]. The algorithm ranks a project for the target user by comparing its similarity to other projects. For example, many users engaging in CoCoRaHS (an outdoor precipitation monitoring project) also sustained engagement in StallCatchers (an online project designed to accelerate Alzheimer's research).

Therefore, Stall Catchers will be recommended to a user who engages in CoCoRaHS if they have not yet already engaged in Stall Catchers. The similarity is calculated by cosine similarity.

3.2.3 Matrix Factorization - SVD

This algorithm predicts the user's rating for each project and projects with the highest predicted ratings are recommended to the user [24]. SVD states that any matrix R can be factorized as: $R = USV^T$. This algorithm is used in recommendation system in order to find the multiplication of the three matrices U, S, V^T, to estimate the original matrix R and hence, to predict the missing values in the matrix. (The matrix R includes missing values since users did not participate in all projects, and therefore we would like to estimate how much will a user like an unseen project). In the settings of recommendation system, the matrix U is a left singular matrix, representing the relationship between users and latent factors. S is a diagonal matrix describing the strength of each latent factor, while V^T is a right singular matrix, indicating the similarity between items and latent factors. Latent factors describe a property or concept that a user or an item have. For example, a latent factor can refer to where the project takes place (indoors or outdoors, online or offline, etc.). SVD decreases the dimension of the utility matrix R by extracting its latent factors. It maps each user and item into a latent space with r dimensions and with this, we can better understand the relationship between users and projects, as they become directly comparable.

For example, the project Asteroid Mappers (an online project designed to identify craters on the asteroid Vesta) is recommended to user A, since this project has the highest predicted rating than all other projects, for this target user. It has the highest predicted rating mainly because the user has participated in the past in projects Moon Mappers (an online project designed to perform science tasks on Mars) and StallCatchers (described above), which are similar to project Asteroid Mappers and also because project Asteroid Mappers best describes the correlation of the target user with other users.

3.2.4 Popularity Based

In this algorithm, all projects are sorted by popularity level [1]. Popularity is measured by the number of users who contribute to the project. For example, the project Globe at night (a project designed to measure the night sky brightness) is one of the most popular projects in SciStarter, and many users participate in it. Therefore, this project is recommended for users assigned to this algorithm.

3.3 OFFLINE SETTING

In this section, we evaluate the different algorithms on historical data from SciStarter. We used data that included logged-in SciStarter-users, who interacted with at least two projects. (Interaction is defined above). The experiment included 6353 users, where each user was assigned to each one of the algorithms.

3.3.1 Evaluation Methodology

We chronologically split the data into train, and test sets such that 10% of the latest interactions from each user are selected for the test set and the remaining 90% of the interactions are used for the train set.

We evaluate the top-n recommendation result. We use two of the most common top-n metrics: Precision and Recall. Precision: the proportion of recommended projects that are relevant (Relevant projects are the projects that the user eventually interacted with, either by clicking and/or participating), Recall: proportions of relevant projects that were recommended. In addition, we also examined a new metric named Refined Precision, which was suggested in [10]: out of the rejected items (recommended but not interacted), Refined Precision describes the sum of max similarities to interacted projects. Precision, Recall and Refined Precision were measured at 3, where 3 is the number of recommended projects. The number 3 was chosen since SciStarter presents three recommended projects by default.

3.4 ONLINE SETTING

After evaluating the algorithms in an offline setting, we also wanted to test their performance online. We performed three online experiments. The first experiment took place between 16.9.2019 to 11.10.2019. The second experiment took place between 11.10.2019 - 2.12.2019 and included recommendations based on user's physical location. In the first experiment, users were provided with recommendations which were not necessarily close to the user physical location. Therefore, it was not possible for users to participate in such projects. For example, the project "Vermont Atlas of Life on iNaturalist" takes place in Vermont, U.S, and users from Europe, could not participate in this project. Hence, we restricted our algorithms to recommend only projects that are physically close to the user physical location. By taking into account the IP address with which the user enters SciStarter, we could infer her landmark, and use it as a restriction for the recommendations. For each such user, we filtered the recommended projects by this restriction and recommended on the top n projects which are in the same region as the user or projects which are taken online. Lastly, the third experiment took place between 2.12.2019 - 17.2.2020. In this experiment, we initialized users allocation to cohorts from fresh, since the clicks stream data which our algorithms were based on, was unreliable before. In this paper, we will describe only the last experiment.

3.4.1 Participants

A total of 1331 logged-in users have participated in the online experiment, which was conducted between December 2019 and February 2020. Out of the 1331 users, only 131 users interacted with at least two projects. These users were assigned equally into five cohorts, while the rest of the users got the default recommendation of the Popularity algorithm. The users allocation to cohorts was round robin allocation since we wanted to cohorts to be in the same size. Each cohort was provided with recommendations based on the different algorithm: (1) Item-based Collaborative filtering, (2) User- based Collaborative filtering, (3) Matrix Factorization - SVD,(4) Popularity based, (5) Baseline. Where 'Baseline' is a cohort which did not get any recommendation from the system and instead, provides the user with Scistarter's promoted projects as an alternative to the recommendations.

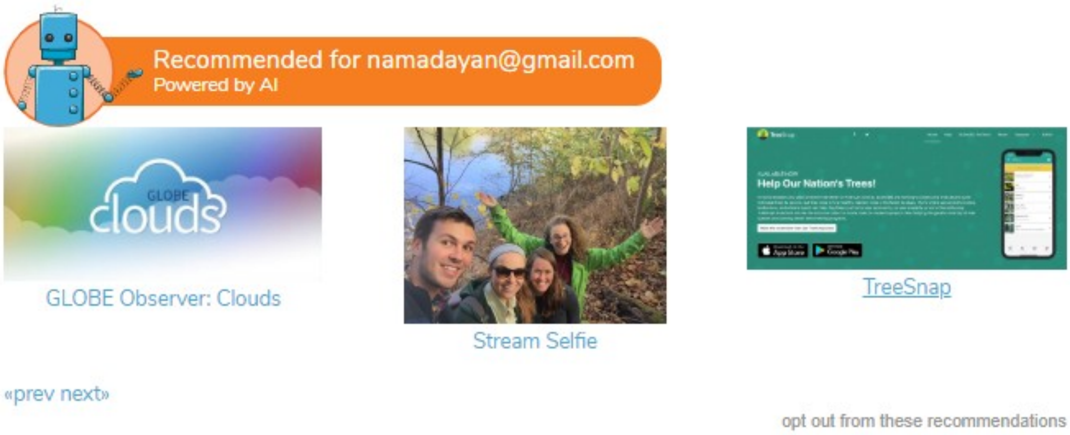

FIGURE 1: SCREENSHOT OF RECOMMENDATION TOOL FROM ONLINE EXPERIMENT

3.4.2 Evaluation Methodology

In our system, we use live evaluation via A-B testing to evaluate the recommendation system performance. Users are assigned to one of the five cohorts described above, where the baseline cohort is the control group. We measure precision, recall and refined precision metrics in the online experiment as well. However, in our domain, the number of recommended items presented to the user changes. Users are presented with three projects and can choose to get more recommendations. Thus, instead of calculating precision, recall and refined precision at 3, where 3 is the number of recommended projects, we calculate these metrics at the actual number of recommendation each user is exposed to.

In order to evaluate recommendation quality in the online experiment, we use a number of different metrics:

- RecE - number of users' actions of clicking on a recommended project and/or interacting with it later.
- NoRecE - number of users' actions where the user did not click on a recommended project but did interacted with the project later.
- NoRecNoE - number of users' interactions with projects that were not recommended to her, and
- ToolE - number of any users' engagement with the recommendation tool (e.g. clicking the project, clicking on the project's image or 'next', 'prev' buttons).

With these metrics, we evaluate system changes on live traffic and also track the system performance at an ongoing basis.

Co-funded by the Horizon 2020
Framework Programme of the European Union

4 RESULTS

4.1 OFFLINE RESULRS

Table 1 contains results of off-line evaluation for the 4 examined algorithms, where the described metrics are precision, recall and refined precision at 3. User-based collaborative filtering and SVD outperform the other algorithms. Popularity algorithm generates the lowest performance on precision and recall.

Algorithm	Precision@3	Recall @3	Refined Precision@3
Item Based Collaborative filtering	4.6%	43.2%	41.7%
User Based Collaborative filtering	22.7%	68.6%	53.7%
Popularity	11.1%	33.3%	44.5%
SVD	18.6%	54.9%	45.0%

TABLE 1: OFF-LINE RESULTS FOR RECOMMENDATIONS ALGORITHM TYPES.

4.2 ONLINE RESULRS

The recommendation tool was active on SciStarter for over 41 days and gave the following results:

26-27 users were allocated to each cohort, and no user has chosen to opt out from the experiment. A total of 624 clicks on recommendations were recorded, which is about 5.6% CTR. (including users with less than 2 interactions). Table 2 presents the described recommendation quality metrics for each examined algorithm. We can see that SVD algorithm has the best performance in all metrics.

Algorithm	#Users	RecE	NoRecE	NoRecNoE	ToolE	Precision	Recall	Refined Precision
Item Based Collaborative filtering	25	13	3	56	25	2.9%	8.8%	25.3%
User Based Collaborative filtering	25	36	6	37	72	11.3%	28.2%	35.4%
Popularity	25	38	1	29	99	9.1%	22.5%	36.7%
SVD	25	54	8	64	124	11.4%	24.4%	37.8%
Baseline	25	15	1	29	29	9.3%	24.2%	37.0%

TABLE 2: ON-LINE RESULTS FOR RECOMMENDATION ALGORITHM TYPES.

Some correlation exists between the offline and the online experiments. In both experiments SVD and User based collaborative filtering were the best algorithms. However, the performance of Item based collaborative filtering algorithm was significantly lower in the online experiment as compared to the offline experiment, while popularity algorithm proves itself better in the online experiment. We can see that 3 out of 4 algorithms (User based collaborative filtering, SVD and popularity) are better than the baseline algorithms in all tested metrics. Users are much more engaged with SciStarter and the projects as compared to the baseline approach.

We can reason SVD's good performance according to the literature, where SVD is considered as a leading algorithm in the domain of recommendation systems [12, 23]. In addition, in this particular work, SVD generates recommendations that are more rare (less familiar to users) than other algorithms do. A user study we have performed has shown that users are more interested in projects they have never heard about: "I did not click on either project because I have looked at both projects (several times) previously", "I am more interested in projects I didn't know exists before". This finding is another reason that explains why SVD performs better than the other algorithms. We can demonstrate the difference in the recommendations of SVD, which recommends more rare projects than other algorithms.

- User A, who is assigned to User based collaborative filtering algorithm receives the recommended projects: [167,734,2992], which their popularity scores are [11,16,19] respectively.
- User B, who is assigned to Item based collaborative filtering algorithm receives the recommended projects: [2992,83,114], which their popularity scores are [19,12,25] respectively.
- User C, who is assigned to SVD algorithm receives the recommended projects: [890,671,87], which their popularity scores are [3,1,6] respectively.

Co-funded by the Horizon 2020
Framework Programme of the European Union

We can see that SVD recommends on less popular projects than other algorithms.

5 USER STUDY

In order to learn what are the users' opinions about the recommendations, and their level of satisfaction, we conducted a survey with SciStarter's users. Our survey was sent to all SciStarter community users.

One hundred and thirty eight users have filled the survey, where each user was asked about the recommendations presented to him by the algorithm he was assigned to. The survey included questions about users' overall satisfaction with the recommendation tool as well as questions about their pattern of behavior before and after the recommendations. See Appendix A for the survey details. The majority of users (75%) were very satisfied with the recommendation tool and claimed that the recommendations were suitable for their personal interests and goals. The majority of users (54%) reported they have clicked on the recommendations and visited the project's site (Figure 2), while only 8.8% of users did not click the recommendation or visited the project site, users who were not familiar with the recommended projects before, clicked more on the recommendations, as well as users who previously performed a contribution to a project.

Users who did not click on the recommendations can be divided into 3 main themes: (1) Users who don't have the time right now or will click the project in the future. (2) Users who feel that the recommendations are not suitable for their skills and materials: "Seemed out of my league", "I didn't have the materials to participate". This behaviour was also discussed in [26], and was named "classification anxiety". (3) Users who feel that the recommendations are not suitable for their interests: "No interest in stall catchers", "The photos and title didn't perfectly match what I am looking for". Not many users have also performed a contribution to the recommended projects (only 15%), but they reported their desire to do so in the future. About 93.4% of users reported they would use the AI recommendations tool in the future, while 92% of them are familiar with the search engine.

The survey demonstrated how powerful the impact of AI is. We got some very enthusiastic feedback on the recommendations: "I am very impressed by the new Artificial Intelligence feature from SciStarter! Your AI feature shows me example projects that I didn't know before exist", "I like how personalized recommendations are made for citizen science users", etc.

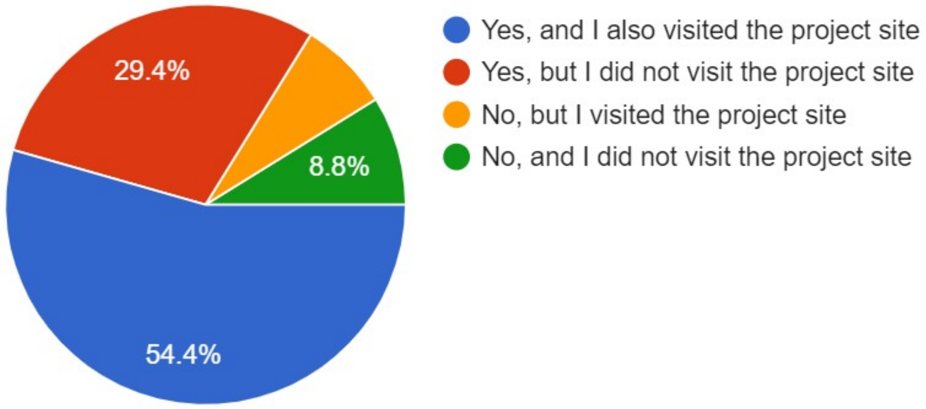

FIGURE 2: DID YOU CLICK ON ONE (OR MORE) OF THE RECOMMENDED PROJECTS?

6 SUPPORT FOR DIVERSITY

The WeNet project focuses on supporting diversity in its developed theories, algorithms and implementations. Thus we wanted to check how are the algorithms developed in this research performing for diverse populations. Identifying gaps in such performance will be the basis for our work in the next phases of the WeNet project.

For this task we inspect the performance of our recommendation algorithms for two diversity dimensions. The first dimension is the location of the user that interacted with the system. For this we separate the users in the system according to their location continent and check their average interaction with the recommendation system. This dimension is a fixed dimension that does not consider the nature of the user interaction with the system. The second dimension is the amount of user interaction with the system prior to the activation of the recommendation system. We divide users to three groups: low interaction, medium interaction and high interaction based on their average number of contributions before the experiment and inspect their response to the recommendation algorithm during the experiment. This is a dynamic dimension that looks at users according to their behaviour and not according to some fixed demographic characteristic.

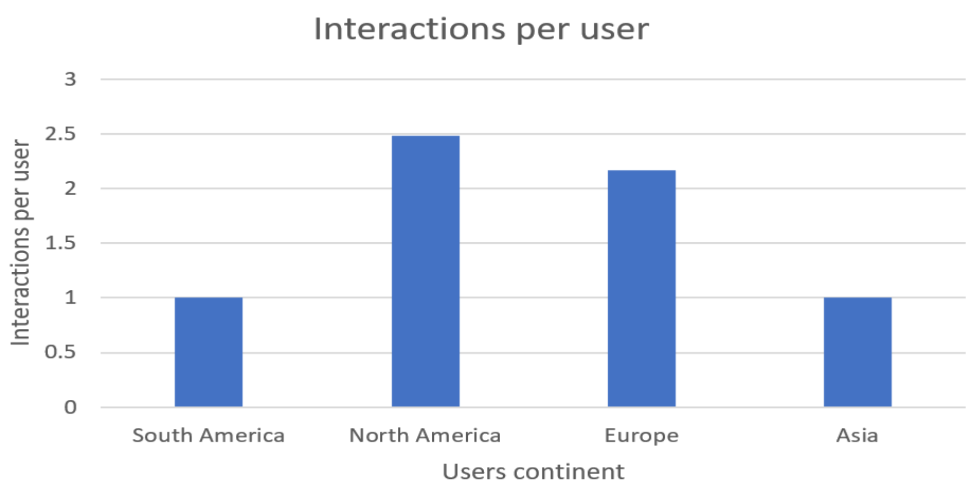

FIGURE 3: INTERACTION PER USER PER CONTINENT

Figure 3 presents the average interaction with the recommendation tool per user divided by continents. As can be seen from the figure, users from North America and Europe had a significantly higher response to the recommendation algorithms than users

Co-funded by the Horizon 2020
Framework Programme of the European Union

in South America and Asia. This points to a diversity gap in the current algorithms as they seem to be less effective / attract less interest from some continents compared to others.

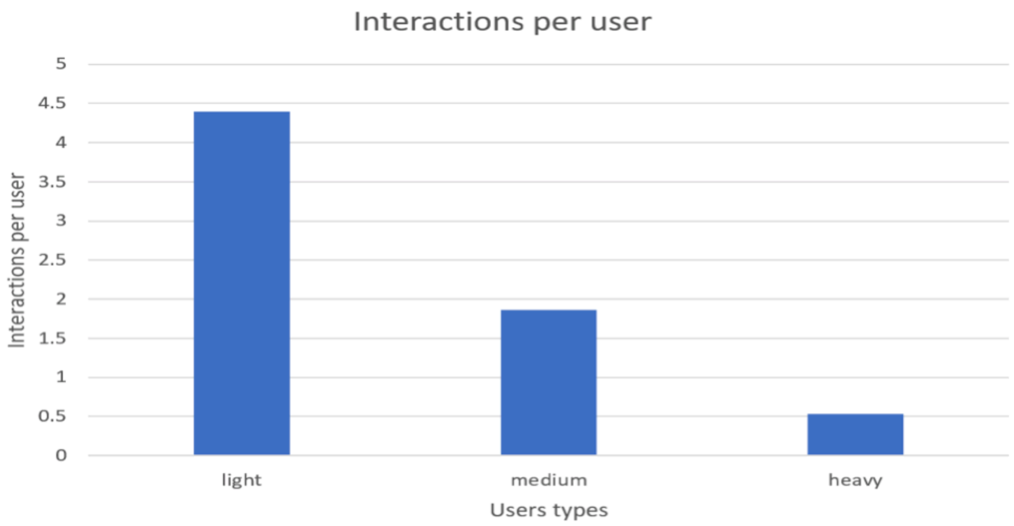

FIGURE 4: INTERACTION PER USER TYPE

Figure 4 presents the average interaction with the recommendation tool per user divided user types. Users are divided to three types based on the amount of their interaction with the system prior to the introduction of the recommendation algorithms. As can be seen from the figure, the current recommendation algorithms have strong impact on light users and are showing minimal impact on heavy users, in terms of interest as demonstrated through the interaction with the recommendation.

We note the above results as indications of the work needed as next steps for this development effort. Specifically, both static user features as well as dynamic user features should be considered when developing the next steps of our algorithms focused on better support for diverse populations.

7 WIDER SIGNIFICANCE AND CONCLUSION

Citizen science engages volunteers in scientific research. The prevalence of the internet has significantly increased the scale of citizen science projects and members dramatically increases. With this wide variety of projects, finding the right project, which best suits the user preferences and capabilities, has become a major challenge and is essential for keeping the user motivated and active.

Our research presents a novel approach for supplying intelligent recommendations to users based on artificial intelligence techniques. We demonstrate this approach in a study conducted in the SciStarter platform and including hundreds of users. We find that a particular recommendation algorithm that is based on collaborative filtering generated significantly more user activity than other algorithms. Collaborative filtering methods predict the fitness of a project for a user based on matching the history of activity of the user with that of other users. Based on the successful results of the project, the recommendation tool we built is being incorporated into the live SciStarter platform.

A crucial part of the study was to quantitatively understand the relationships between SciStarter users and project activities (including project page views, bookmarks, joins, and contributions). We used a combination of data including SciStarter's clickstream, users' profiles, project meta data, to explore patterns among participants and the projects they engage with. Among the 3,000+ projects registered on SciStarter, we narrowed our research to 80 SciStarter affiliate projects because these projects use APIs (see Participant API documentation: SciStarter.org/API) that report participants' contributions back to SciStarter and back to the user. We limited the scope of research to SciStarter account holders (vs visitors to the site). Among the 70,000 SciStarter accounts, we further narrowed the study participants to those who engaged in at least two projects.

We use machine learning to explore the relationships between a user's activities and profile on SciStarter, their referral source (where they came from before they reached SciStarter.org), their communities (Girl Scouts, schools, etc), and others who viewed, saved, joined or contributed to sim- ilar projects. Users were divided into cohorts. Then, once software detected patterns of engagement, algorithms were tested to select and display only three recommended projects to the select cohorts. Our project demonstrated that AI could successfully be deployed to personalize task recommendations and help would-be participants more easily discover the task most suitable for them. The outcome of this research project is the AI-powered Recommendation Widget currently displayed to logged in users of SciStarter. This is visible on a user's homepage and dashboard.

The results of our research have been featured on the DiscoverMagazine.com blog, on the SciStarter podcast, on the SciStarter.org blog, and on SciStarter's social media. Approximately 70,000 SciStarter users (including thousands of project leaders) have been informed about the project and encouraged to login to SciStarter to test the new feature. Future plans for the AI-powered Recommendation Widget include integration with a new SciStarter project to extend its offerings to include ALL forms of public engagement in science.

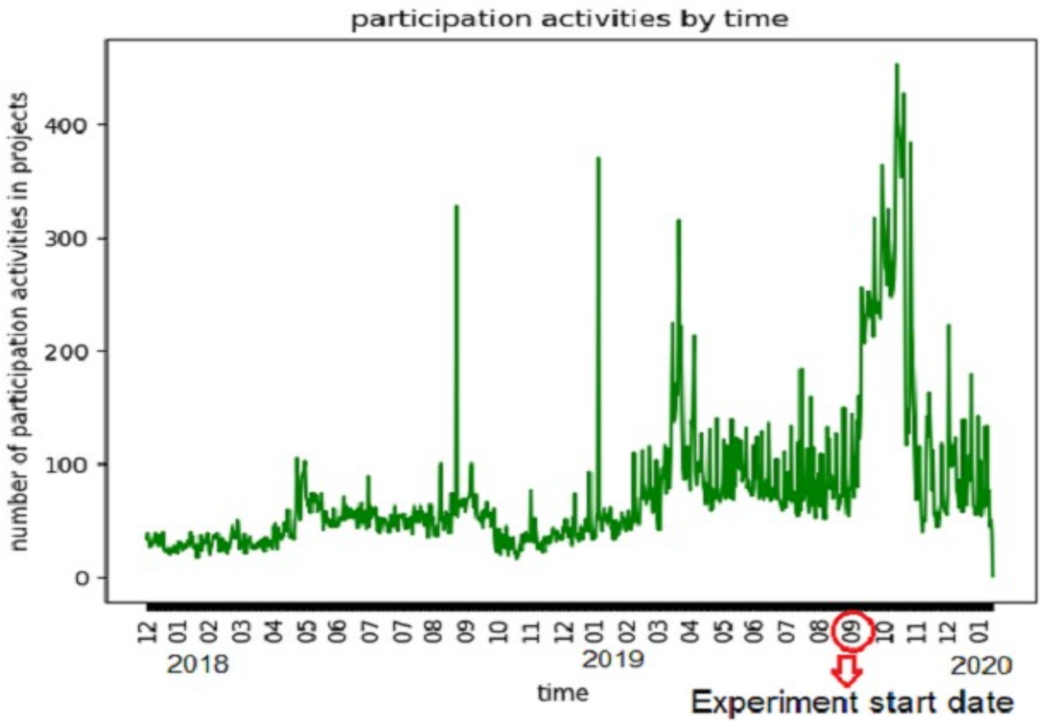

FIGURE 5: PARTICIPATION ACTIVITY BY TIME

This project has transformed how SciStarter helps projects recruit and support participants and better respond to their needs. It was so successful in increasing engagement, that SciStarter has decided to make the widget a permanent feature of their site. This will help support deeper, sustained engagement to increase the collective intelligence capacity of projects and generate improved scientific, learning, and other outcomes.

The use of the recommendation algorithm that we have developed has already made a substantial impact on SciStarter. Figure 5 demonstrates the increase in user participation in projects. Since experiment start date, there is a significant increase in the number of users participation activities.

In addition to the direct impact on SciStarter, our project demonstrated a generalized approach for improving collective intelligence in citizen science by connecting users, data and AI. It showed that artificial intelligence can be used to guide citizen scientists through the SciStarter eco-system, matching them with projects selected by other users with similar characteristics, based on their profiles and logged activities. It directly increased user engagement to enhance science learning and improve crowdsourced data quality.

Our future work directly builds on the results of this project to develop improved algorithms, which support diversified user populations. As such, these algorithms will receive as input various social characteristics of user groups and will be able to better recommend tasks and actions to users based on their diversified preferences and charactersitics while preserving privacy and fairness.

When considering the different algorithms tested in this work, we note that user based collaborative filtering and SVD outperformed the other options tested. As such, these two algorithms will be the first to be tested in the context of the WeNet platform for recommendation based incentivizing. Our future algorithms should also adhere to transparency requirements as described in out latest work on transparency for machine generated personalization [33]. As such, these algorithms should encompass the ability to explain to the user what was the reasoning behind the recommendations generated. We note that in this aspect, user based collaborative filtering offers a straight forward explainability approach (based on similarity) while SVD explainability requires additional investigation to be carried out in future work (due to its reliance on latent factors in the user-item space).

Additionally, we plan to combine our recommendation based approach with badge allocation, and compare two badge path- ways through SciStarter, a platform with 3,000 projects and 65,000 registered users, This combination will enable us the build a recommendation and badge based platform that will be also applicable for the WeNet platform when ready.

We will measure two type of badge approaches:

1) completion- based: recognition for completion of online tasks to measure acquired competencies and contributions to projects; and

2) quality-based: recognition quality of participation.

Badges for the first pathway, designed for quantity, will embed a list of tasks that participants performed, and address social problems of motivating for participation. Applications will provide visible badge icons, and information about how to achieve the different badges. Badges for the second pathway, designed for quality, will use social feedback to acknowledge performed task quality and contribution to the wider community.

We will explore how badges can support sustained engagement in applications as well as improve the quality of users contributions - by engaging stakeholders in the design and testing of these pathways using assessments, analytics, clickstreams and surveys to evaluate preferences and project data quality.

Our hypotheses is that (1) badges will encourage sustained engagement benefitting a) participants through formalized recognition, and b) application builders through more and better data; (2) the optimal badge pathway will vary based on communities and motivations and should include (as in the case of recommendations) a robust diversity aware mechanism.

REFERENCES

[1] H. J. Ahn. Utilizing popularity characteristics for product recommendation. International Journal of Electronic Commerce, 11(2):59–80, 2006.

[2] M. Aristeidou, E. Scanlon, and M. Sharples. Profiles of engagement in online communities of citizen science participation. Computers in Human Behavior, 74:246–256, 2017.

[3] M. Balabanović and Y. Shoham. Fab: content-based, collaborative recommendation. Communications of the ACM, 40(3):66–72, 1997.

[4] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk. Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11):977–984, 2009.

[5] D. Brossard, B. Lewenstein, and R. Bonney. Scientific knowledge and attitude change: The impact of a citizen science project. International Journal of Science Education, 27(9):1099–1121,2005.

[6] H. K. Burgess, L. DeBey, H. Froehlich, N. Schmidt, E. J. Theobald, A. K. Ettinger, J. HilleRis- Lambers, J. Tewksbury, and J. K. Parrish. The science of citizen science: Exploring barriers to use as a primary research tool. Biological Conservation, 208:113–120, 2017.

[7] S. Dwivedi and V. K. Roshni. Recommender system for big data in education. In 2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH), pages 1–4. IEEE, 2017.

[8] D. M. Fleder and K. Hosanagar. Recommender systems and their impact on sales diversity. In Proceedings of the 8th ACM conference on Electronic commerce, pages 192–199. ACM, 2007.

[9] J. Freyne, M. Jacovi, I. Guy, and W. Geyer. Increasing engagement through early recommender intervention. In Proceedings of the third ACM conference on Recommender systems, pages 85–92. ACM, 2009.

[10] S. Frumerman, G. Shani, B. Shapira, and O. Sar Shalom. Are all rejected recommendations equally bad?: Towards analysing rejected recommendations. In Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, pages 157–165. ACM, 2019.

[11] C. Funk, J. Gottfried, and A. Mitchell. Science news and information today. Pew Research Center, 2017.

[12] S. Gower. Netflix prize and svd, 2014.

[13] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In 2008 Eighth IEEE International Conference on Data Mining, pages 263–272. Ieee, 2008.

Co-funded by the Horizon 2020
Framework Programme of the European Union

[14] J. Itmazi and M. Gea. The recommendation systems: Types, domains and the ability us- age in learning management system. In Proceedings of the International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[15] G. Kragh. The motivations of volunteers in citizen science. environmental SCIENTIST, 25(2): 32–35, 2016.

[16] H. Labarthe, F. Bouchet, R. Bachelet, and K. Yacef. Does a peer recommender foster students' engagement in moocs?. International Educational Data Mining Society, 2016.

[17] J. Laut, F. Cappa, O. Nov, and M. Porfiri. Increasing citizen science contribution using a virtual peer. Journal of the Association for Information Science and Technology, 68(3):583–593, 2017.

[18] O. Nov, O. Arazy, and D. Anderson. Scientists@ home: what drives the quantity and quality of online citizen science participation? PloS one, 9(4):e90375, 2014.

[19] L. Ponciano and T. E. Pereira. Characterising volunteers' task execution patterns across projects on multi-project citizen science platforms. In Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems, page 16. ACM, 2019.

[20] M. J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S. Szalay, and J. Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. arXiv preprint arXiv:0909.2925, 2009.

[21] F. Ricci, L. Rokach, and B. Shapira. Introduction to recommender systems handbook. In Recommender systems handbook, pages 1–35. Springer, 2011.

[22] F. Ricci, L. Rokach, and B. Shapira. Recommender systems: introduction and challenges. In Recommender systems handbook, pages 1–34. Springer, 2015.

[23] R. A. Sadek. Svd based image processing applications: state of the art, contributions and research challenges. arXiv preprint arXiv:1211.7102, 2012.

[24] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Incremental singular value decomposition algorithms for highly scalable recommender systems. In Fifth international conference on computer and information science, volume 1. Citeseer, 2002.

[25] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative filtering recommender systems. In The adaptive web, pages 291–324. Springer, 2007.

[26] A. Segal, Y. Gal, R. J. Simpson, V. Victoria Homsy, M. Hartswood, K. R. Page, and M. Jirotka.
Improving productivity in citizen science through controlled intervention. In Proceedings of the 24th International Conference on World Wide Web, pages 331–337, 2015.

[27] A. Segal, K. Gal, E. Kamar, E. Horvitz, and G. Miller. Optimizing interventions via offline policy evaluation: Studies in citizen science. In Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[28] Q. Wu, H. Wang, L. Hong, and Y. Shi. Returning is believing: Optimizing long-term user engagement in recommender systems. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pages 1927–1936. ACM, 2017.

[29] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer. 2009 Aug 7;42(8):30-7.

[30] Drachsler H, Verbert K, Santos OC, Manouselis N. Panorama of recommender systems to support learning. InRecommender systems handbook 2015 (pp. 421-451). Springer, Boston, MA.

[31] Pathak B, Garfinkel R, Gopal RD, Venkatesan R, Yin F. Empirical analysis of the impact of recommender systems on sales. Journal of Management Information Systems. 2010 Oct 1;27(2):159-88.

[32] Gomez-Uribe CA, Hunt N. The netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS). 2015 Dec 28;6(4):1-9.

[33] Schelenz L, Segal A, Gal K. Best Practices for Transparency in Machine Generated Personalization. arXiv preprint arXiv:2004.00935. 2020 Apr 2.

APPENDIX A: QUESTIONNAIRE

Hello there, SciStarter community member!

We have recently released a tool providing you with personalized project recommendations. This tool uses AI technology to provide you with new project ideas that are best suitable for your enjoyment! You can see an example of the tool below.

You can help us out to evaluate the effectiveness of the recommendation tool. We will hand out $10 Amazon gift cards to the first 100 responders . You can read more about the recommendation tool at this link To be eligible for the gift card please do the following:
1. Log on to SciStarter with your username and take notice of your personalized recommendations.
2. Fill out the survey below when you are done with your SciStarter session.

----

1. Were you familiar with any of the projects that were recommended for you? *

○ Yes
○ No

2. Did you click to find out more information on one (or more) of the recommended projects? *

○ Yes, and I also visited the project site
○ Yes, but I decided not to visit the project site No, but I intend to visit the project site
○ Later
○ No, and I do not intend to visit the project site

3. If you did NOT want to click on any of the recommended projects, can you tell us why?

....

4. If you decided to visit any of the recommended projects, did you also make a contribution to that project?

○ Yes
○ No

5. How satisfied are you from the recommendation tool (1-Not satisfied at all; 5 – Very satisfied)? *

1	2	3	4	5

Co-funded by the Horizon 2020 Framework Programme of the European Union

6. How suitable were the recommended projects for your personal interests and goals (1- not suitable at all; 5 – Very suitable)? *

1	2	3	4	5

7. Do you plan to use the AI recommendation tool in the future?

- ○ Yes
- ○ No

8. What are your main reasons for choosing to contribute to a project? *

- ○ The topic of the project interests me (topic could be physics, birds, insects,
- ○ education, etc.) The task of the project interests me (photography, classification, transcription, etc.)
- ○ The project is nearby
- ○ The project can be done online
- ○ The project is similar to projects I contributed to in the past
- ○ The project is well known and popular
- ○ Other

9. How many total projects have you contributed to in SciStarter? *

- ○ I didn't contribute to any project
- ○ I contributed to 1-2 projects
- ○ I contributed to more than 2 projects

10. Have you ever used the SciStarter project search tool? *

- ○ Yes
- ○ No

11. Please let us know about any other comments and ideas you have regarding project recommendations in SciStarter. Your ideas are appreciated!

....
....
....
....

12. What gender do you identify as? *

Co-funded by the Horizon 2020
Framework Programme of the European Union

○ Male
○ Female
○ Prefer not to say

13. What is your age? *

○ 0-15 years old
○ 16-30 years old
○ 31-45 years old
○ 46-60 years old
○ 61+

14. What is the highest degree or level of education you have completed? *

○ In middle school
○ In high school
○ High school degree or equivalent Bachelor's degree (e.g. BA, BS)
○ Master's degree (e.g. MA, MS, MEd)
○ Professional degree (e.g. MD, DDS, DVM) Doctorate (e.g. PhD, EdD)
○ Prefer not to say

15. Please check the box to confirm that you filled out the survey and are eligible for $10 Amazon gift card *

○ I confirm

Thank you! The first 100 respondents are eligible to receive a $10 Amazon gift card via email. Please enter your email address below so we can send you the card if you match! To protect your anonymity, we cannot make any connection between your email and your survey.

....

Co-funded by the Horizon 2020
Framework Programme of the European Union