# Investigating possible solutions for the cold start problem in a food recommender application

Fundamenten van Mens-machine interactie [G0Q55a]

Felix Cammaerts
Nathan Dedoncker
Jules Hermans
Timo Martens
Tom Martens
r0663453
r0588064
r0601036
r0703070
r0630037

## ABSTRACT

Recommender systems are being deployed in a lot of different domains. Companies like Spotify, Amazon and Netflix use it to recommend items to their customers. However, recommender systems often suffer from the cold start problem. This problem occurs when the system doesn't know anything about the preferences of the user. This can happen when a new user logs in for the first time. To address this problem we developed a food recommender application. Two possible solutions for the cold start problem are implemented in the application. An online, within-subjects user study was conducted. Each participant filled in an online survey after interacting with three variants of the application. The user was asked to either do nothing, rate meals or to select meals he/she would like to cook as an initialization process. The recommender system then used these preferences to provide recommendations. Qualitative and quantitative evaluation methods were applied to gather results. Both solutions seem to offer more accurate and personalized results. The participants also felt that the recommended meals were similar to each other. The needed effort and time are also important aspects of the application. The results show that the users don't mind the extra effort, if better recommendations are produced.

## 1 INTRODUCTION

Rise of the planet of the *apps*. No, this paper will not be a review of the first part of a movie franchise about apes[1]. Instead, this will be a discussion on recommender applications. The increasing interest in apps for our smartphones is shaping our daily lives. It is no longer only geeky developers that are playing around with these, but also the average layman. As a result of this, user interface design is an important academic field. This paper will be an analysis of recommender apps in particular. Big technology companies such as Netflix, Spotify, etc. are using your data to suggest new movies or playlists. There is one issue with this though - what if you just started using their app and they have not gathered any information on you yet? In research, this is often called the *cold start problem.* We will try to tackle this problem using our own recommender app called Foodversity. Its main purpose is to propose new and exciting recipes based on what you liked in the past. We will start this study off by looking at related work in section 2. After this, in section 3 an explanation of Foodversity and its implementation follows. Next up, in section 4 an experiment will be introduced in which the following research questions will be posed:

- How do 3 different registration procedures compare when evaluating the quality of generated recommendations and do these procedures solve the cold start problem?
- What do users think about the user-friendliness of these various procedures?

An experiment is useless without results, so these can be found in section 5. We will close out this paper with a discussion of these results in section 6.

## 2 RELATED WORK

The inspiration for this research comes from our own weekly routines as students. We need to compose new meal plans every week and coming up with healthy and tasty recipes every week is not an uncomplicated task. We often resort to typical student meals, which are quick to cook and generally unhealthy. This is the ground on which this application is built. Hence, finding a diverse set of recipes based on things that a user already likes is the core principle of Foodversity.

As to our knowledge, there are few apps that share the same formula as Foodversity. One of them is Yummly [2]. They have a clean design with a straightforward user interface, but there are some weaknesses we believe. A first weakness is a good initialization procedure. Registration is very easy thanks to integration with popular platforms such as Google and Facebook, yet the user is never asked for his/her preferences such as certain diets. This seems to be the reason for many complaints. Looking at their interface, we can also assume that there is no easy way to rate recipes that you've tried. Therefore it looks like *your feed* is solely based on recipes

[1]See planet of the apes

[2]The app can be found at https://play.google.com/store/apps/details?id=com.yummly.android

that are viewed regularly and is not personalized. From these 2 shortcomings, we can conclude that there is definitely some room for improvement.

Cold start is an important problem when working with recommender applications. Since cooking takes a significant amount of effort, if a certain recipe isn't to your taste it can feel like a waste of man's most important resource, time. So it is crucial to suggest relevant recipes for our users. Starting off with wrong suggestions might also lead to users deleting your app as it feels ineffective. That is why cold start is the main research topic of this paper. The research done by Elahi et al [2] provides a great foundation to start from. As we will be applying a recommendation mechanism based on collaborative filtering, we can use their proposed techniques to deal with cold start. Two of these approaches are elicit rating and elicit preference in which the recommender system will exploit the user to gain some starting data. How these are applied precisely is explained in section 4.

## 3 IMPLEMENTATION

### 3.1 Data

Life is simpler with API's, isn't it? Well, that is exactly what we thought as well. This is why Foodversity employs the Spoonacular API [3]. Their database consists of over 5000 meals and contains images, nutritional information and full recipes with cooking material and ingredients that are needed. Also available are features such as searching for similar recipes and searching based on certain diets e.g. vegetarianism or veganism. By using an API, we are able to omit the search for a data set and can be more complete because we have access to a hefty amount of recipes.

### 3.2 Recommender system

As explained in section 2, we need a recommender algorithm based on collaborative filtering. Which means that we would be using the ratings of previous meals to predict the likelihood of a user fancying a new recipe. If we do this, we can apply elicit rating and elicit preference as cold start mitigating techniques. We designed our own recommender mechanism which combines user preferences and recipe similarity. Whenever the user rates a certain recipe, it is saved. Every time that he requests new suggestions, Spoonacular is used to acquire recipes that are similar to the rated meals. These will then be ordered on the rating of the original recipe that they are similar to. This means that a recipe which is comparable to 2 different recipes will get their combined ratings, making it rank higher. Pseudocode on how the recommendation scores get updated for a user is provided in algorithm 1.

### 3.3 Interface

The first thing to do when designing the user interface of an application is to know your target audience and what exactly the app would be used for. This is where personas and storyboards come in. A persona is a fictional character that represents the goals and behavior of a hypothetical user group. A storyboard is a visualized scenario in which a user would be using the application. Using these visuals, our team obtained a common view on what Foodversity should look like. The next step in the prescribed design methodology is a low fidelity prototype of the application. This is used to have an initial design and involve the end user a first time. It is made in Google slides which enables us to connect screens with the use of hyperlinks. Testing this design is done by conveying a think aloud study with 7 participants of which the demographics are given in table 6. The users were given 5 tasks to complete which are given in table 7. This left us with a couple of improvement points, mostly relating to the user interface design itself and not the user-friendliness of the application. We changed these remarks when developing a high-fidelity prototype.

Once the initial high-fidelity prototype was created, we conducted another think aloud study to ensure there were no problems left with the interface. 7 new participants were recruited of which the demographics are given in table 8. We asked the participants to do the 9 tasks listed in table 9. Due to the fact that some of these tasks are very similar (i.e. when a user has completed one of the tasks, the other one becomes trivial) we created Latin squares for these tasks. This way each task gets put in front of every other task an equal number of times. These Latin squares were made for the tasks 5 & 6 and 7, 8 & 9, their respective Latin squares can be found in table 10 and 11. Results of the study did not show any big issues, so this implied that the design of the app is clear. We also asked participants to complete a post-test survey taken from SUS [1]. This resulted in a score of 84.6 and as can be seen in figure 3, this means that Foodversity scores excellent on system usability.

The high fidelity prototype and final application are written using the Flutter framework[4]. The most important screen in our application is the recommendation screen (figure 1), this is where the users get to see which recipes are being recommended to them. This screen has changed a lot after its initial conceiving in the low-fidelity prototype. For instance, there were no yellow info buttons yet. These allow the user to get extra details about the recipe's nutritional values. Another issue that we discovered thanks to the feedback from the think aloud study was that it was not clear how many recipes a user should select. Henceforth we changed the button at the bottom to "Select *n* meals" in order to make it clear that the user can select however many meals they prefer.

It should be noted that the application is slightly adapted to perform the testing. The registration procedure now has the option to choose between 2 recommender initializations. In the top right corner of the recommendation screen, there used to be a button to filter the suggestions. This was not needed to do our research, so we replaced it with a number that represents the total amount of generated recommendations. Using this number, we can evaluate the quality of the proposed recipes.

[3]The documentation for this API can be found at https://spoonacular.com/food-api/docs

[4]Information can be found at https://flutter.dev/


**Figure 1: Recommendation screen**

# 4 EXPERIMENT

## 4.1 Participants

As the main motivation for this application is having students eat healthier, the partakers of our study are primarily scholars themselves. They are recruited through our generation's most popular means of communication, social media. It is crucial that participants are not closely related to the researchers, as that would make them biased. Only people that are merely acquaintances or former research partners - who fully understand the importance of being neutral - are approached. The demographics of the test population can be found in table 12. As far as how many participants, we followed the common rule: "The more partakers, the more accurate our results will be". We ended up with a total of 38 participants.

## 4.2 Experimental design

We will be comparing 3 different registration procedures, which will produce both quantitative and qualitative results. We opted for a within-users study as this allows us to compare the results between the different initialization procedures more easily. One essential thing to keep in mind with such studies is the occurrence of learning effects or fatigue. This means that a participant might perform more efficient on the second registration procedure because they know what's coming. The remedy for this is counterbalancing, which means changing the order in which they experience the various registration procedures.

The application is available on the Google play store[5]. Participants are guided through this experiment via a questionnaire, which is added in appendix B. Working with a questionnaire provides us with a standardized testing routine. Note that the participants are not abandoned by us. We are always available to them if they have any questions regarding the survey, but it remains important to not assist them with their usage of the application.

## 4.3 Measurements

First, they are asked some general questions about their behaviour. Following that is the first registration procedure. This one is very simple and doesn't try to fix the cold start problem. It simply generates random recipes. The user is then asked to count the meals that he would enjoy out of these. He is also asked to give his opinion on this extremely basic user experience. When this is done, the user is asked to use the application and go through 2 different initializations of the recommender system. One of these is elicit preference in which the user selects his favourite meals out of a randomized list. The recommender system then uses this as a basis for new suggestions. The second initialization is elicit rating in which the user rates random recipes, these are then used as a foundation for the second recommender mechanism. For both procedures the user is asked the same questions as for the simple registration. We now have both quantitative and qualitative results for the 3 procedures. Using these, the methods can be compared against each other in both quality of the recommendations and user experience. The results will be discussed in the following sections.

To evaluate the 3 procedures, we formulated a set of questions based on the ResQue questionnaire by Li et al [3]. A total of 12 questions were composed which are split up in terms of accuracy, novelty, user satisfaction, diversity and effort. Seeing that users did not have to complete a registration in the first process, the questions concerning effort were omitted for process 1. The questions are as follows:

- The items recommended to me match my interests (ACC1).

[5] It can be found under this address https://play.google.com/store/apps/details?id=jules.hermans.fmmi

- This recommender system gave me good suggestions (ACC2).
- The items recommended to me were unexpected (NVY1).
- The items recommended to me were pleasantly surprising (NVY2).
- The items recommended to me were helpful (SAT1).
- I would cook some of the items recommended to me (SAT2).
- the recommendations did not feel personalized (DIV1).
- I was provided with a lot of similar recipes (DIV2).
- I feel that the registration process was not complicated (EFF1).
- I feel that the registration process is efficient (EFF2).
- I feel that the registration process is clear (EFF3).
- I feel that the registration process takes too much effort/time (EFF4).

The answers to these question are in the format of a five-level Likert scale. This means the answers range from 1 (strongly disagree) to 5 (strongly agree). Additionally, some open questions were formulated, such as stating how much meals the user wants to cook, given his or her recommendations.

# 5 RESULTS

## 5.1 Statistics

The data that was collected via the user surveys consists of both quantitative (ratio) and qualitative (ordinal) data. After each initialization procedure the user was asked to fill in the amount of meals he/she would like to cook given her recommendations, but the number of actual recommendations differed for each individual. Therefore, we transformed this absolute number (of meals selected) to a ratio (RAT). To analyze the answers between the different initialization processes, Wilcoxon signed rank tests were used for both data types. In both cases the independent variables were the initialization processes whereas dependent variables were users' answers. Process 2 (P2) is the initialization process where the user had to select meals he/she would like to cook. Process 3 (P3) consists of rating the meals as the initialization process. And process 1 (P1) has no initialization procedure. As stated before, we would like to see how the 3 different processes compare to one another from a users' perspective.

## 5.2 Quantitative results

The results of the Wilcoxon test on the ratio of selected meals are shown in tables 1 and 2. From these tables it is possible to conclude that both initialization methods 2 and 3 are an improvement for the cold start problem. A pairwise comparison between P1 and P2 reveals that P2 is a better solution for the cold start problem ($p<0.005$), since the ratio of selected meals is higher. The same conclusion is made from the comparison of P1 and P3 ($p<0.001$). Between P2 and P3 there was no significant difference in ratios.

**Table 1: Result of Wilcoxon signed rank test (for ratio data) showing significant differences between process 1 and 2. SD = standard deviation**

| | P1 | | P2 | | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Z | P |
| RAT | 0.435 | 0.25 | 0.571 | 0.23 | -2.957 | 0.003 |

**Table 2: Result of Wilcoxon signed rank test (for ratio data) showing significant differences between process 1 and 3. SD = standard deviation**

| | P1 | | P3 | | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Z | P |
| RAT | 0.435 | 0.95 | 0.639 | 0.22 | -4.142 | <0.001 |

## 5.3 Qualitative results

The tables included in this section only contain the significant difference between the corresponding processes. The insignificant results are left out for the purpose of illustrating clear tables. Table 3 contains the pairwise comparison between P1 and P2. The survey asked the users to rate each process in terms of different aspects using a Likert scale. Table 3 shows that P2 scores significantly better in terms of Accuracy ($p<0.003$). P1 is better at recommending new items to the user ($p<0.03$). And finally, the users feel that P2 recommends more personalized and similar recipes ($p<0.03$). Table 4 reveals the comparison between P1 and P3. A similar conclusion as with table 3 can be made. P3 shows significant improvement in terms of Accuracy and Diversity ($p<0.001$). The meals recommended by P1 are also more unexpected ($p<0.001$). The difference is that P3 also scores significantly better for Satisfaction ($p<0.03$). Finally a comparison is made between P2 and P3 in table 5. This reveals that P3 offers better suggestions ($p<0.05$) and the suggestions are perceived as more personalized ($p<0.02$). Figure 2 illustrates an overview of the average answers for each question of our questionnaire. This graph emphasises the improvement for the cold start problem for both P2 and P3 in comparison with P1. It also shows that the participants didn't mind going through the initialization process to achieve better recommendations. The effort of each process was measured using four questions about effort and clarity of the initialization procedure.

**Table 3: Result of Wilcoxon signed rank test (for ordinal data) showing significant differences between process 1 and 2. M = mean, SD = standard deviation**

| | P1 | | | P2 | | | | |
|---|---|---|---|---|---|---|---|---|
| | M | Median | SD | M | Median | SD | Z | P |
| ACC1 | 2.816 | 3 | 0.95 | 3.684 | 4 | 0.81 | -3.515 | <0.001 |
| ACC2 | 2.921 | 3 | 1.00 | 3.526 | 4 | 0.92 | -3.132 | 0.002 |
| NVY1 | 3.737 | 4 | 0.86 | 3.026 | 3 | 0.91 | -3.25 | 0.001 |
| DIV1 | 3.079 | 3 | 1.19 | 2.605 | 2 | 0.89 | -2.184 | 0.029 |
| DIV2 | 2.711 | 2.5 | 0.96 | 3.447 | 3.5 | 1.01 | -2.872 | 0.004 |

# 6 DISCUSSION

The previous section concludes that both initialization processes (P2 and P3) are suitable for solving the cold start problem. Both procedures have a significant higher ratio of selected meals in comparison with P1. Also, it's clear that both processes offer personalized, accurate and similar results. These were desirable outcomes, as this was the main goal of the initialization process. The tastes of the users are gathered through an initialization procedure. These preferences are then used to suggest similar meals. This explains why

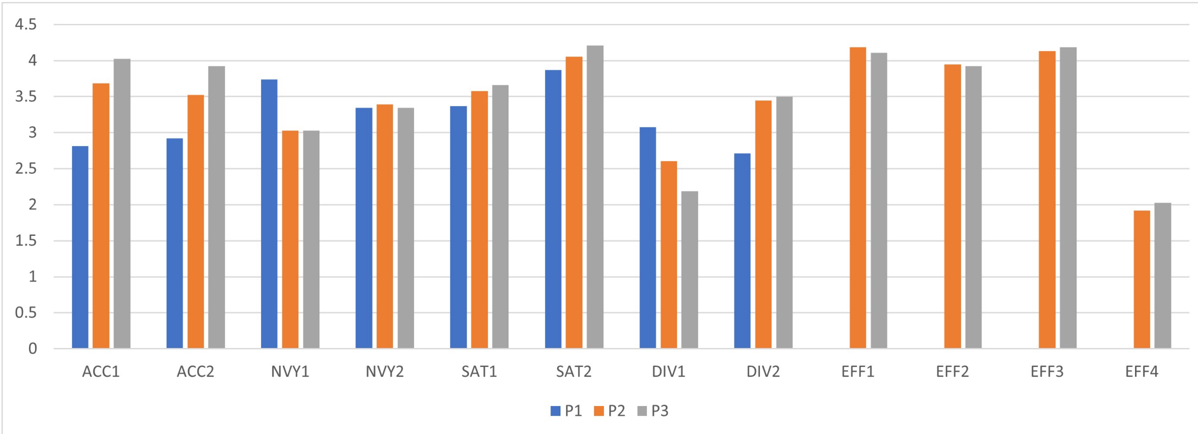


**Figure 2: Results of the questionnaire**

the recommended items feel accurate, personalized and similar. The final lesson learned is that the participants preferred going through a setup to achieve more accurate and personalized recommendations over receiving less accurate or personalized recommendations without going through an initialization. They didn't mind the extra effort or time, because they were offered better recommendations. It is however difficult to select one best procedure, because the data only shows a significant difference in accuracy and diversity. P3 offers better and more personalized recommendations.

## 7 CONCLUSION

In this research we developed a food recommender application for students. It helps students in deciding what to cook by taking into account their preferences. A common problem of such an application is: how to recommend items if you don't know the preferences of the users yet. We tried tackling this problem by implementing an initialization process which polled for those interests. This information is then used to suggest accurate and personalized recommendations. Two different initialization procedures were tested and compared with the system without initialization. Both strategies have a positive effect on the cold start problem. The users felt that the recommendations were more personalized and accurate. These results might not be desirable if the effort or time to achieve them was too high. To find out if this was the case, we included questions measuring these aspects. From this paper we can conclude that the extra needed effort was worth it if they received better recommendations. We expect that an initialization will almost always lead to better and more personalized recommendations. This is probably applicable in other domains as well, however further research is needed to confirm these expectations.

## REFERENCES

[1] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (1995).

[2] Mehdi Elahi, Francesco Ricci, and Neil Rubens. 2016. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review* 20 (2016).

[3] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. 157–164.

**Table 4: Result of Wilcoxon signed rank test (for ordinal data) showing significant differences between process 1 and 3. M = mean, SD = standard deviation**

| | P1 | | | P3 | | | | |
|---|---|---|---|---|---|---|---|---|
| | M | Median | SD | M | Median | SD | Z | P |
| ACC1 | 2.816 | 3 | 0.95 | 4.026 | 4 | 0.59 | -4.437 | <0.001 |
| ACC2 | 2.921 | 3 | 1.00 | 3.921 | 4 | 0.59 | -4.193 | <0.001 |
| NVY1 | 3.737 | 4 | 0.86 | 3.026 | 3 | 0.97 | -3.733 | <0.001 |
| SAT1 | 3.368 | 3.5 | 0.79 | 3.658 | 4 | 0.63 | -2.295 | 0.022 |
| SAT2 | 3.868 | 4 | 0.74 | 4.211 | 4 | 0.62 | -2.595 | 0.009 |
| DIV1 | 3.079 | 3 | 1.19 | 2.184 | 2 | 0.87 | -3.377 | 0.001 |
| DIV2 | 2.711 | 2.5 | 0.96 | 3.500 | 4 | 0.83 | -3.756 | <0.001 |

**Table 5: Result of Wilcoxon signed rank test (for ordinal data) showing significant differences between process 2 and 3. M = mean, SD = standard deviation**

| | P2 | | | P3 | | | | |
|---|---|---|---|---|---|---|---|---|
| | M | Median | SD | M | Median | SD | Z | P |
| ACC2 | 3.526 | 4 | 0.92 | 3.921 | 4 | 0.59 | -2.156 | 0.031 |
| DIV1 | 2.605 | 2 | 0.89 | 2.184 | 2 | 0.87 | -2.537 | 0.011 |

# A APPENDIX A: RESEARCH METHODS

## A.1 Recommender system

**Algorithm 1:** Algorithm to update the recommendation scores for a user.

```
Result: Updates the recommendations for user i
for each ratedRecipe of user i do
    /* Get similar recipes from API */
    similarRecipes = getSimilarRecipes(ratedRecipe);
    for each similarRecipe in similarRecipes do
        /* Retrieve the rating of the rated recipe */
        rating = getRating(ratedRecipe);
        if hasScore(i, similarRecipe) then
            /* If the user already has a recommendation score for this recipe,
               add the new rating to the score. */
            addToScore(i, similarRecipe, rating);
        else
            /* If the user does not have a recommendation score for this
               recipe, create one. */
            addNewScore(i, similarRecipe, rating);
        end
    end
end
```

## A.2 Think-aloud study low-fidelity prototype

**Table 6: Demographics of the participants**

| Age | Job | Amount of times person cooks per week | Gender |
|---|---|---|---|
| 22 | Student | 1 | Male |
| 22 | Student | 3 | Male |
| 22 | Student | 5 | Male |
| 24 | Student | 4 | Male |
| 24 | Student | 5 | Female |
| 20 | Student | 3-4 | Female |
| 21 | Student | 1 | Male |

**Table 7: The tasks that a participant is asked to perform**

| Task number | Task |
|---|---|
| 1 | Create an account |
| 2 | Log in using an existing account |
| 3 | Choose the recipes you would like to cook this week |
| 4 | Go through the steps of preparing any meal you like |
| 5 | Add any ingredient you like to the shopping list |

## A.3 Think-aloud study high-fidelity prototype

**Table 8: Demographics of the participants**

| Age | Job | Amount of times person cooks per week | Gender |
|---|---|---|---|
| 21 | Student | 7 | Female |
| 20 | Student | 7 | Female |
| 20 | Student | 4 | Female |
| 25 | Student | 3 | Female |
| 30 | Sales | 3 | Male |
| 24 | Student | 7 | Female |
| 23 | Student | 5 | Male |

**Table 9: The tasks that a participant is asked to perform**

| Task number | Task |
|---|---|
| 1 | Create an account |
| 2 | Log in using an existing account |
| 3 | Choose the recipes you would like to cook this week |
| 4 | Go through the steps of preparing any meal you like |
| 5 | Add any ingredient you like to the shopping list |
| 6 | Remove any ingredient from the shopping list |
| 7 | Change the amount of store visits |
| 8 | Change your allergy to peanuts |
| 9 | Tell the application that you are following a vegan diet |

**Table 10: Latin square for tasks 5 & 6**

| Order | Task | Task |
|---|---|---|
| 1 | 5 | 6 |
| 2 | 6 | 5 |

**Table 11: Latin square for tasks 7, 8 & 9**

| Order | Task | Task | Task |
|---|---|---|---|
| 1 | 7 | 8 | 9 |
| 2 | 8 | 9 | 7 |
| 3 | 9 | 7 | 8 |

**Table 12: Demographics of final user study**

| Age | Job | Gender | Nationality | Tech savvy |
|---|---|---|---|---|
| 23 | Student | Female | Belgian | Average |
| 24 | Student | Male | Belgian | Average |
| 22 | Student | Male | Belgian | Above average |
| 24 | Engineer | Male | Belgian | Average |
| 23 | Project Manager | Male | Belgian | Above average |
| 23 | Student - athlete | Male | Belgian | Above average |
| 21 | Student | Male | Belgian | Average |
| 21 | Student | Male | Belgian | Advanced |
| 24 | Student | Male | Belgian | Above average |
| 27 | Teacher | Male | Belgian | Average |
| 24 | Student | Male | Belgian | Average |
| 23 | Student | Male | Belgian | Advanced |
| 19 | student | Male | Belgian | Advanced |
| 24 | Projectmedewerker | Male | Belgian | Average |
| 20 | Student | Male | Belgian | Advanced |
| 26 | Audioloog | Female | Belgian | Average |
| 23 | Student | Male | Belgian | Average |
| 29 | Optician | Female | Belgian | Average |
| 18 | Student | Female | Belgian | Beginner |
| 26 | Sales Manager | Male | Belgian | Above average |
| 21 | nurse | Female | Belgian | Above average |
| 21 | Psychology student | Female | Hungarian | Average |
| 23 | Project Engineer | Male | Belgian | Average |
| 22 | Student | Male | German | Advanced |
| 22 | Student | Male | Belgian | Average |
| 60 | Engineer | Male | Belgian | Advanced |
| 22 | Student | Prefer not to say | Belgian | Advanced |
| 21 | Student | Male | Belgian | Above average |
| 20 | Student | Female | Belgian | Average |
| 22 | Student | Male | Belgian | Advanced |
| 21 | Student | Female | Belgian | Average |
| 23 | UX Design Researcher | Male | Belgian | Advanced |
| 22 | student | Male | Belgian | Above average |
| 26 | Government official | Female | Belgian | Above average |
| 23 | Engineer | Male | British | Average |
| 23 | Student | Male | Belgian | Average |
| 19 | Student | Male | Belgian | Above average |
| 24 | student | Male | Belgian | Above average |

# B APPENDIX B: QUESTIONNAIRE

In this questionnaire we will ask you some questions regarding our application and your experience with it. First we will ask some demographic information and afterwards three different versions will be shown and you will be asked to answer some questions about each version If anything is not clear, please contact your contact inside our research team.

## B.1 Demographics

(1) First name
(2) Choose your gender
- Female
- Male
- Prefer not to say
- Other: ....

(3) How old are you?
(4) What is your nationality?
(5) In what country do you currently live?
(6) What is your profession?
(7) Please select your education level
- Primary school
- High school
- College
- Graduate school

(8) How would you rate yourself as a computer user?
- No experience
- Beginner
- Average
- Above average
- Advanced

(9) Do you trust a person/thing even though you have very little knowledge of it?
- Absolutely not
- Probably not
- Probably
- Very probably
- Definitely

## B.2 First version

(10) Based on the list in Figures **??** and 4, which contains random recommended meals for you, how many would you select that you would like to cook? (fill in the amount).

*B.2.1 Accuracy.* Select in the following statements the best suited answer

(11) The items recommend to me match my interests
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(12) This recommender system gave me good suggestions?
- Strongly disagree
- Disagree
- Neutral
- Agree

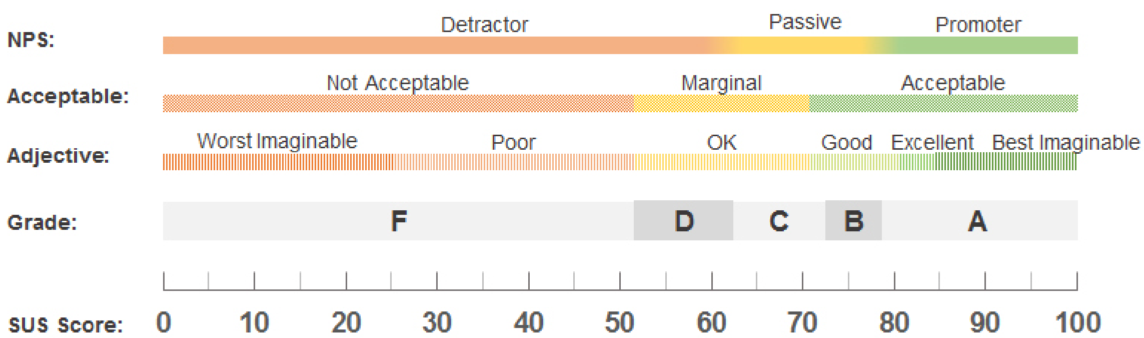


**Figure 3: System usability scale**

- Strongly agree

*B.2.2 Novelty.* Select in the following statements the best suited answer

(13) The items recommended to me where unexpected
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(14) The items to me were pleasantly surprising
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.2.3 Satisfaction.* Select in the following statements the best suited answer

(15) The items recommended to me were helpful
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(16) I would cook some of the items recommended to me
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.2.4 Diversity.* Select in the following statements the best suited answer

(17) I was only provided with general recommendations (e.g. top rated meals) which are the same for anyone (i.e. the recommendations did not feel personalized)
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(18) I was provided with a lot of similar recipes
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

## B.3 Second Version

Before you start this part of the questionnaire, we ask you to go through the registration process of the app that can be downloaded via https://play.google.com/store/apps/details?id=jules.hermans.fmmi and select "option 1" in the registration process after that complete this section of the questionnaire.

(19) You can click on "select your meals" button in the home screen after your are registered (it might take some time for the meals to load). Based on this list you see on this page, how many meals would you select that you would like to cook? (fill in the amount) At the moment there is a bug in the app, please count your selected meals manually and do not trust the counter at the bottom.

(20) On the same screen you should normally see a number in the right hand corner (next to "what meals would you like to cook this week"). What number is displayed?

*B.3.1 Accuracy.* Select in the following statements the best suited answer

(21) The items recommend to me match my interests
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(22) This recommender system gave me good suggestions?
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.3.2 Novelty.* Select in the following statements the best suited answer

(23) The items recommended to me where unexpected
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(24) The items to me were pleasantly surprising
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.3.3 Satisfaction.* Select in the following statements the best suited answer

(25) The items recommended to me were helpful
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(26) I would cook some of the items recommended to me
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.3.4 Diversity.* Select in the following statements the best suited answer

(27) I was only provided with general recommendations (e.g. top rated meals) which are the same for anyone (i.e. the recommendations did not feel personalized)
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(28) I was provided with a lot of similar recipes
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.3.5 Effort.* The effort of the registration process is measured, this is from the moment of clicking "register" until receiving the "account created successfully" screen is considered.

(29) I feel that the registration process was not complicated
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(30) I feel that the registration process is efficient
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(31) I feel that the registration process is clear
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(32) I feel that the registration takes too much effort/time
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(33) If you agreed, or strongly agreed with the previous statement, could you specify why you thought the registration process took to much effort/time? Which elements/functionality took to much effort?where could effort be reduced?

## B.4 Third Version

Before you start this part of the questionnaire, we ask you to first click on log out in the right top corner on the home screen. Then go through the registration process again, but select option 2 instead of option 1.

(34) You can click on "select your meals" button in the home screen after your are registered (it might take some time for the meals to load). Based on this list you see on this page, how many meals would you select that you would like to cook? (fill in the amount) At the moment there is a bug in the app, please count your selected meals manually and do not trust the counter at the bottom.

(35) On the same screen you should normally see a number in the right hand corner (next to "what meals would you like to cook this week"). What number is displayed?

*B.4.1 Accuracy.* Select in the following statements the best suited answer

(36) The items recommend to me match my interests
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(37) This recommender system gave me good suggestions?
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.4.2 Novelty.* Select in the following statements the best suited answer

(38) The items recommended to me where unexpected
- Strongly disagree

- Disagree
- Neutral
- Agree
- Strongly agree

(39) The items to me were pleasantly surprising
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.4.3 Satisfaction.* Select in the following statements the best suited answer

(40) The items recommended to me were helpful
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(41) I would cook some of the items recommended to me
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.4.4 Diversity.* Select in the following statements the best suited answer

(42) I was only provided with general recommendations (e.g. top rated meals) which are the same for anyone (i.e. the recommendations did not feel personalized)
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(43) I was provided with a lot of similar recipes
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

*B.4.5 Effort.* The effort of the registration process is measured, this is from the moment of clicking "register" until receiving the "account created successfully" screen is considered.

(44) I feel that the registration process was not complicated
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(45) I feel that the registration process is efficient
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(46) I feel that the registration process is clear
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(47) I feel that the registration takes too much effort/time
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(48) If you agreed, or strongly agreed with the previous statement, could you specify why you thought the registration process took to much effort/time? Which elements/functionality took to much effort?where could effort be reduced?


**Figure 4: List of recommended meals**