Mohammed V University Rabat

National Advanced School of Computer Science and System Analysis, Morocco

Master of Big Data & Data Science

MASTER THESIS PROJECT

Designing and Developing a Personalized Country Recommender System

Supervisors:

BEN HIBA Lamia, ENSIAS

EL IONI Nabil, UNIBZ

ELAHI Mehdi, UNIBZ

Student:

EL MAJJODI Ayoub

October, 2019

Abstract

Nowadays, many people decide to change their current country and to choose a new place to immigrate to. A big challenge in this decision is how to find the best possible destinations that would best match the particular needs and constrains of people, e.g., their professional background or their expectations in terms of life conditions and facilities. Although there are a number of world ranking lists that annually compare and publish the best places to live, none of them offer personalized ranking list that fit users preferences.

The goal of this thesis is to address this issue by designing, implementing and evaluating a personalised Recommendation System of countries. The system is capable of eliciting preferences of users, learning from the preferences, and intelligently generating a personalised ranking list of countries for each target user. The quality of the system has been evaluated in an offline and real user study, and measured in terms of accuracy, diversity, novelty, satisfaction and capability to understand the particular preferences of different users.

Acknowledgements

First of all I would like warmly thank my supervisors, Pr. Lamia Benhiba, Pr. Nabil EL Ioini and Pr. Mehdi Elahi, for the support, confident they show me, and inspiration and relevant advice at all levels.

I express my heartfelt thanks to the members of the jury who honored me with their presence. I would also like to thank all the teachers who contributed my training during my years of study and in particular, the two years of study in Master Data Science and Big Data at ENSIAS.

Finally, this thesis was made possible thanks to my family's encouragement. I would like to thank my parents, brothers and sisters for their support throughout my years of study, without them, nothing would have been possible.

Praise be to God

to my family

to the sons of my tribe

those who didn't find

the opportunity to pursue

their studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the last decades, a tremendous amount of data has become available in the World Wide Web, that makes users often encounter difficulties in finding the right product or appropriate information that better suits diverse personal interest and needs [3]. Most users try to seek assistance from others who had previously had the same needs for those particular information or items. Recommender Systems (RSs) can tackle this problem by guiding users through the process of finding the right digital content and product based on user interests and taste. Such tool are revolutionizing the web and by their effective support to the users. Accordingly, it seems that *"we are leaving the age of information and entering the age of recommendation" [4].*

RSs found success in e-commerce and entertainment domain as they present the items or goods that are likely to be of interest of the user and hence increase user satisfaction and engagement. Lately RSs have been increasingly used in other fields such as travel [1] and tourism [5, 6], restoration [7], health [8, 9], education [10, 11] and even e-business domain [12, 13].

This thesis addresses a different application domain of RSs, i.e., immigration. These days, people move around a lot and due to variety of reasons, prefer to start a new life in new countries all over the world. According to statistics there were around 244 million international migrants in the world in 2015 [2], this large number generated, by human endeavors to enhance either quality of life, professional condition or just satisfying specific needs. Migration affects beneficially economic dimension and go beyond to other dimensions of human development such as education, health, social development,

[1] TripAdvisor,Expedia.com,Visitfinland.com

[2] UN DESA, 2016

and the transfer of skills.

The main challenge that faces people who tend to undertake a journey to another country, is to identify the possibly *best* destination that would best match their professional and personal profiles and their expectation in term of life satisfaction and facilities. There are a large variety and volume of options (e.g., the number of cities in the world), that are published annually presenting world ranking lists of the best places to live [14, 15]. However, none of these lists is personalised according to individual needs and constrains of people. However, a personalised system with a recommendation of countries (or even cities) can tackle this challenge, and providing a *usable* tool which ultimately assist the people to better make decision on where to live. This is indeed one of the most important decisions to make and certainly need expert supports (that can be automatic or manual).

The aim of this thesis is to address this issue by designing, implementing and evaluating a personalised RSs for countries and cities. The system initially build user profiles by eliciting their preferences (in different forms), and learning from preferences and ultimately generate a personalised ranking list of countries and cities for each target user based on her preferences.

In order to achieve our objective, we formulated a number of research questions :

- **RQ1 :** Which recommender algorithms can be adopted -based on the preferences of users in order to generate personalized country ranking ?
- **RQ2 :** What are the most important features that users consider when deciding to move to another country ?
- **RQ3 :** Do recommender algorithm preferences depend on personality types ?
- **RQ4 :** Will the system for generating personalized country ranking be usable according to the user's assessment ?

The remainder of this thesis is organized as follows. In Chapter 2, we present the state of the art regarding RSs, in order to well understand the different approaches of RSs, how they are used as a method to predict user preferences and make recommendations, and how they are evaluated. Then we briefly discuss the relationship between human decision making and personality. Next in Chapter 3 we describe the methodology followed in this project, the design and the implementation details. The results analysis of the experiments is given in Chapter 4 with the discussion of important observations. Finally, in Chapter 5 we elaborate conclusions and give future work directions.

Chapter 2

State of the Art

Introduction

This chapter serves to establishes the key concepts and vocabulary for the rest of the master thesis, we define different approaches of recommender system, recommender systems evaluation. At the end we will briefly discuss the relation between human decisions making the personality.

2.1 Recommender Systems

First of all it is necessary to understand what Recommender Systems (RSs) is, what type of RS exists, and what components are included.

Recommender Systems emerged as an independent field of research in mid-1990's and derived from different other areas, such as cognitive science, information retrieval, approximation theory, and also have a link to management science [16].
Resnick et Varian [17] describe the underworkings of a recommendation systems system as :

> *"People provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients "*(Resnick and Varian, 1997)

This seems to be a suitable definition for the early recommender systems. since then RSs has greatly benefited from combined interest and effort that industry and academia have given to this field. Nowadays the recommender system term can be described in a wider way :

> *"Any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options."*(Burkee 2002) [18]

Recommender system is able to provide suggestions to different users, in multiple contexts, Mayer et al [19] identifies 4 key features for any RS:

- Help to decide : Given item i, user wants to know if he will appreciate it.
- Help to compare : Given n items, user wants to know what item to choose
- Help to explore : Given item i , user wants to know what are the related k items
- Help to discover : Given a huge catalogue or items, user u wants to find k new interesting items

According to the above definitions, that the main goal of every RS guiding the user through the process of figuring out suitable objects, items, information and goods that suit their needs, RSs go so far as to finding what user hasn't even thought about yet, that they may want in the future.

2.2 Data and Knowledge Sources

Recommender systems are information processing systems that actively gather various kinds of data in order to build their recommendations.
In most cases there are recommendation techniques that are knowledge poor, i.e, they use sample and basic data, like user ratings/evaluations, other technique seeks for complex knowledge such as, ontological description of user or items, social relations or user activities etc.
In almost any case, as a general categorization the data used by RSs, can be divided into three kinds of objects [3] :

Items: Items are objects that are recommended, can be characterized by their complexity and their value, that can be a positive value if the item is useful for the user, or negative one if the item is not appropriate and the user made a wrong decision when selecting it.

Users: In order to make recommendations more personalized, RSs exploits a range of information about the users, the selection of what information to model depends in most cases on the recommendation technique.

Transaction: Recorded interactions between the user and the RS. Transactions are log-like data that store important information generated human-computer interaction.
In fact, ratings are the most popular form of transaction data an RS collects, these ratings may be collected either, relying on the explicit technique, on the implicit technique or the hybrid one. [20]

- The Explicit technique : process that involves user assigning, numerical or scores rating to evaluate the items or goods retrieved to use by the system.
- Implicit technique : operates and makes decision based on user action, this technique measure user's taste and interest without seeking for user consent.
- Hybrid technique : is the combination of both implicit and explicit techniques, uses both the combinations of numerical rating scores and human behaviour in predicting items of interest and taste of a particular user.

According to [21] collected ratings can be:

(a) Numerical ratings : such as 1-5 stars provided in amazon book recommender.

(b) Ordinal ratings : such as "strongly agree, agree, neutral, disagree, strongly dis- agree" in which the user asked to choose the term that indicates his or her opinion about an item.

(c) Binary ratings : in this type of ratings the user in a sample way asked to decide if a particular item is good or bad.

(d) Unary ratings : can indicate that a user has observed or purchased an item, or otherwise rated the item positively. In such cases, the absence of a rating indicates that we have no information relating the user to the item (perhaps she purchased the item somewhere else).

2.3 Recomendation problem formulation

The general recommendation problem can be described as a problem relying on leveraging user data to render possible future user predictions (likes and interests).
More formally, this verbal problem can be formulated as follows:
Let U be the set of all users, and I the set of all the possible items. The space I can be huge, ranging in hundred of thousand or even millions of items in some applications, like book recommendations, music recommendation, goods recommendation and so on, also the user space U can be large.
Let f be the utility function that measures the suitability of item i to the user u needs, i.e , $f : UxI \rightarrow R$, with R is a totally ordered set. Then for each user $u \in U$, we want to choose such item $i' \in I$ that maximizes the user's utility function, mathematically :

$$\forall u \in U, i'_I = argmax f(u, i)$$

In Recommender systems, the utility of an item is usually represented by a ratings (2.2) that indicates how much a user liked a particular item.

2.4 Recommender systems approaches

In general to this date, there are several approaches to classify recommendation systems, according to how the recommendations are made [22–26]. These are, collaborative filtering [27, 28], content-based filtering [29, 30] and hybrid one [31, 32], that includes both collaborative and content-based filtering.

2.4.1 Collaborative filtering

According to L. Herlocker et al [33] collaborative filtering (CF) algorithms aim to predict a user's affinity for items or information. The filtering decision in CF based on human and not machine analysis of content, each user of an CF system rates items, that they have experienced, in order to establish a user profile of interest. The CF system after that matches together that user with people of similar tastes or interest, then ratings from those similar people are used to generate recommendations for new user. Formally, the utility function (2.3), $f(u, i)$ of item i for user u, is estimated of the utilities $f(u_j, i)$ assigned to item i by those users $u_j \in U$ who are similar to user u.

For example, in a movie recommendation application in order to recommend movies to user u, the collaboration recommender system tries to find the "peers" of user u, i.e., other users that have identical tastes in movies. Then only the movies that are most liked by the "peers" of user would be recommended.

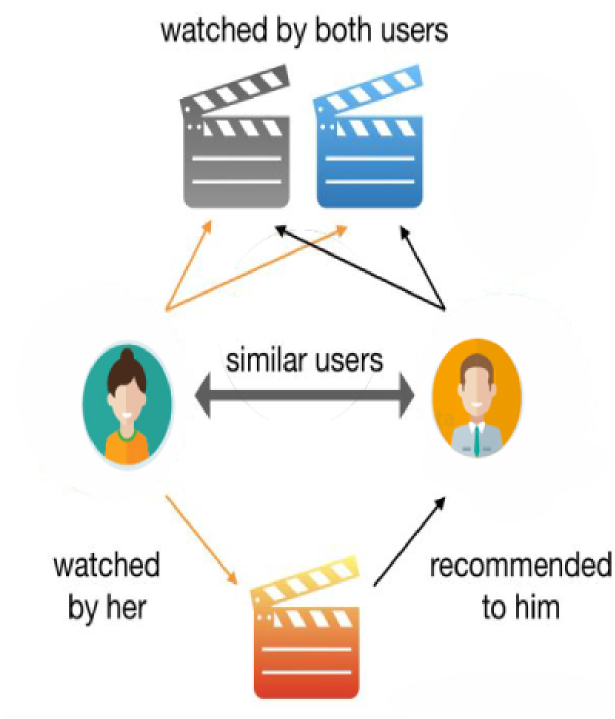


Figure 2.1: CF, movie recomendation example

Moreover CF approach, can be categorized into two classes [34]:

2.4.1.1 Memory based Collaborative Filtering

Memory based CF operates over the entire user database to make predictions, by obtaining similar relationship between users or items according to the user-item rating matrix and then recommending the items that are highly rated by similar users for the active user. In fact, memory-based CF method can be subdivided into, user-user based CF and item-item based CF.

- user-user based CF algorithm: user-user based algorithm, predict a recommendation to active user by processing the following steps:

(1) Calculate the similarity between users, can be computed based on one of the classical measures or others [34].

Cosine similarity: the user's rating can be indicated as n-dimensional vector $r_u = \{r_{u_1}, r_{u_2}, ..., r_{u_n}\}$, and similarity between users is obtained through the user's rating vector angle, the smallest the angle is, the higher the similarity is. Cosine vector similarity is calculated as follows :

$$\begin{aligned} sim_{u,v} &= cos(\vec{r_u}, \vec{r_v}) \\ &= \frac{\vec{r_u}.\vec{r_v}}{\|\vec{r_u}\|_2 \times \|\vec{r_v}\|_2} \\ &= \frac{\sum_{i \in I_{uv}} r_{ui}.r_{vi}}{\sqrt{\sum_{i \in I_u} r_{ui}^2}\sqrt{\sum_{i \in I_v} r_{vi}^2}} \end{aligned}$$

where sim_{uv} represents the similarity between users u and v, $\vec{r_u}$,$\vec{r_v}$ represent the ratings vector of u and v, $\|\vec{r_u}\|$, $\|\vec{r_v}\|$ are 2-norm of u and v, respectively, r_{ui} and r_{vi} represent the ratings of u and v on the item i, I_u and I_v are sets of items rated by users u and v respectively, and I_{uv} represents the set of items commonly rated by both u and v .

Pearson correlation coefficient is calculated as follows :

$$sim_{uv} = \frac{\sum_{i \in I_{uv}} (r_{ui} - \overline{r}_u)(r_{vi} - \overline{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \overline{r}_u)}\sqrt{\sum_{i \in I_{uv}} (r_{vi} - \overline{r}_v)}}$$

where $\overline{r}_u$ and $\overline{r}_v$ represent the average rating from u and v respectively.

(2) Find the Nearest Neighbors, based on two methods, K-nearest neighbors method that select the first k users with the closest similarity to the active user u as his or her nearest neighbors, and the threshold method, that set initially a threshold δ, and when the similarity between user v and current user u is greater than δ, the user v is selected as one of the nearest neighbors.

(3) Predict Ratings, there are two main ways to make recommendations for an active user, predicting the ratings and providing top-N recommendation list. The both need to find rating of the active user u on a new item i using the ratings on i from users most similar to u, the predicted rating can be computed as follows :

$$\hat{r}_{ui} = \overline{r}_u + \frac{\sum_{v \in N_u} sim_{u,v}(r_{vi} - \overline{r}_v)}{\sum_{v \in N_u} |sim_{uv}|}$$

where N_u denotes the similar neighbor set of the user u.

Example :

Table 2.1: Example of movies ratings matrix (on a 5-star scale)

	Jocker	Forrest gump	The professor and the Madam	the Shawshank Redemption
User A	4	?	3	5
User B	?	4	5	?
User C	5	4	2	?
User D	2	4	?	3
User E	3	4	5	?

Consider the ratings matrix above, we want to find user C's prediction for "the Shawshank Redemption" $\hat{r}_{c,shaw}$, with the following configuration :

i) Pearson correlation.

ii) Neighborhood size of 2.

C's mean ratings is 3.66. There are only two users for the neighborhood : A and D. $sim_{(C,A)} = 0.832$ and $sim_{(C,D)} = -0.515$, the prediction $\hat{r}_{C,shaw}$, is therefore computed as follows:

$$\begin{aligned}
\hat{r}_{C,shaw} &= \overline{r}_C + \frac{sim_{(C,A)}(r_{A,show} - \overline{r}_A) + sim_{(C,D)}(r_{D,show} - \overline{r}_D)}{|sim_{(C,A)}| + |sim_{(C,D)}|} \\
&= 3.66 + \frac{0.832.(5-4) + -0.515.(2-3)}{0.832 + 0.515} \\
&= 4.766
\end{aligned}$$

- item-item based CF algorithm : user-user CF, while effective, suffers from scalability problem, as the user base grows, searching for the neighbors of a user is an $O(|U|)$ operation [35]. item-item CF, is one of the most widely used collaborative filtering technique used today, it uses similarities between the ratings patterns of items. Following the same steps of user-user CF, item-item CF can be executed as follows:

  (1) Calculate the Similarity between Items, The classical measures between are :

  Adjusted Cosine vectors, this measures can be calculated as follows :

$$sim_{ij} = \frac{\sum_{u \in U_{ij}} (r_{ui} - \overline{r}_u)(r_{uj} - \overline{r}_u)}{\sqrt{\sum_{u \in U_i} (r_{ui} - \overline{r}_u)^2} \sqrt{\sum_{u \in U_j} (r_{ui} - \overline{r}_u)^2}}$$

where sim_{ij} denotes the similarity between items i and j. U_i and U_j represent the sets of users who rated items i and j, respectively, and U_{ij} denotes the set of users who rated both items i and j.

Pearson correlation coefficient method is calculated as follows:

$$sim_{ij} = \frac{\sum_{u \in U_{ij}} (r_{ui} - \overline{r}_u)(r_{uj} - \overline{r}_u)}{\sqrt{\sum_{u \in U_{ij}} (r_{ui} - \overline{r}_u)^2} \sqrt{\sum_{u \in U_{ij}} (r_{ui} - \overline{r}_u)^2}}$$

where $\overline{r}_i$ and $\overline{r}_j$ represent the average ratings on i and j respectively.

(2) Find the Nearest Neighbors, similar to user based CF, there are usually two methods for finding nearest neighbors in the item-based CF, k-nearest neighbors and setting threshold.

(3) Predict Ratings, item-based CF, generates predictions as follows :

$$\hat{r}_{ui} = \overline{r}_i + \frac{\sum_{j \in U_i} sim_{ij} \times (r_{uj} - \overline{r}_j)}{\sum_{j \in N_i |sim_{ij}|}}$$

where N_i is the similar neighbor set of the item i.

2.4.1.2 Model-based Collaborative Filtering :

Model-based CF, in contrast, uses the user database to estimate or learn a model, which is then used for prediction. The model-based CF, requires a learning phase in advance for finding out the optimal model parameters before making a recommendation, after the learning phase is finished, the model based RS, easily predict the ratings of the active user. A widely adopted method in model-based approach is latent factor model (LFM), which factorizes the user-item rating matrix into two low-rank matrices: the user features and item features matrices. Singular value decomposition (SVD), matrix factorization (MF), non-negative matrix-factorization (NMF) are most usually used recommendation methods among (LFM) method [34], In addition, model-based approach takes advantage of machine learning algorithms, such as Bayesian network, clustering and rule based approaches, in order to process the model building well [36]

2.4.2 Content-based filtering

Content-based (CB) approaches, or thematic filtering, originate from the field of Information Retrieval (IR), the methods known as CB filtering recommends an item to a user while relying upon the description of both the item and the user profile in term of item characteristics, these characteristics are called metadata.

In Content based filtering, the utility function $f(u, i)$ of item i for user u is estimated based on the utilities $u(u, i_j)$ assigned by user u to items $i_j \in I$ that are similar to item i. In other words finding the similar items to the ones that the user liked before. For example, in order to recommend movies to user u, the CB recommender system tries to understand the commonalities among movies that user u has highly rated in the past (specific actor, directors, genres, subject, etc), then only the movies that have high degree of similarity user's preferences would get recommended (see [37]).

In general, the recommendation process in CB method, is performed in three steps depicted in Figure 2.2, each of which is handled by a separate component:

- Content Analyser : when information has no structure, some kind of pre-processing steps is needed to extract structured, meaningful information, the main responsibility of the component is to represent information coming from the source, in a suitable form for the next steps.

- Profile learner : that responsible to collect a representative data of user preferences, and tries to generalize this data, in order to construct the user profile, usually relies on machine learning techniques [38], due to creating a model of user's preferences from the user history is a form of classification learning, the training data of a classification learner is divided into categories, e.g., the binary categories "items the user likes" and "items the user doesn't like"

- Filtering component : this step mainly, exploit the user profile to suggest relevant items, by matching the profile representation against that of items to be recommended.

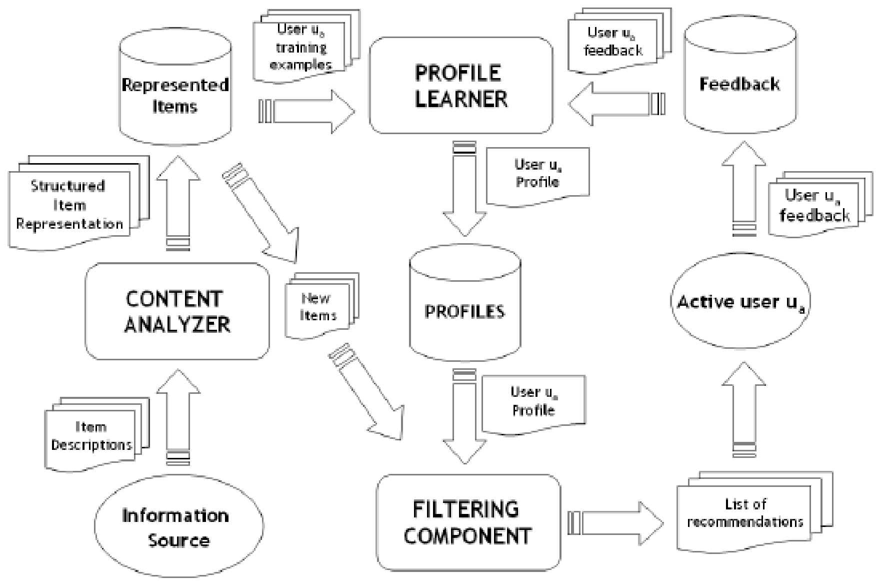


Figure 2.2: Content-based Recommender Architecture [1].

2.4.3 Hybrid filtering

Hybrid recommendation system, is a system based on different data source. Several RSs uses a hybrid approach by combining two or more recommendation techniques, mostly collaborative filtering (CF) and content-based (CB) techniques, to overcame content-based method limitations [39], such as limited content analysis, over-specialization and Cold Start problem [40], and collaborative approach limitations such as new user/item limitations and sparsity etc [16].
On the other hand, the combination of collaborative filtering and content-based methods, can be implemented in one of the following ways :

(1) Implementing collaborative and content-based methods separately, then combining their predictions.

(2) Incorporating some content-based characteristics into collaborative approach.

(3) Incorporating some collaborative characteristics into content-based approach.

(4) Constructing a general unifying model that incorporate both content-based and collaborative characteristics.

In addition, other hybridization methods described in table bellow [18].

Table 2.2: Hybridization Methods

Hybridization method	the Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Mixed	Recommendations from several different recommenders are presented at the same time
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm.
Cascade	One recommender refines the recommendations given by another.
Feature augmentation	Output from one technique is used as an input feature to another.
Meta-level	The model learned by one recommender is used as input to another.

The fact that, basic recommendation techniques have complementary advantages and disadvantages, provided incentive for research in hybrid recommender systems in order to improve the performance ,make accurate recommendations and tackle the main recommendation challenges such as the cold start problem.

2.5 Evaluation of Recommender Systems

The problem of developing good metrics to measure effectiveness of recommendation, has been addressed in many recommendation systems literature [20], [35], [41]. Evaluation of recommendation systems algorithms, is usually done in terms of statistical accuracy of the predicted values with respect to the held-out ratings in a natural starting point for evaluating recommender output. Mean absolute error (MAE) among the statistical accuracy, that measures the average absolute deviation between a predicted rating and the user's true ratings:

$$MAE = \frac{1}{|R|}\Sigma_{(u,i,r)\in R}|\hat{r}_{u,i} - r_{u,i}| \qquad \text{(eq:2.5.1)}$$

In addition, Root Mean Squared Error (RMSE) between the predicted values and actual ratings, can be classified as statistical accuracy measures:

$$RMSE = \sqrt{\frac{1}{|R|}\Sigma_{(u,i,r)\in R}(\hat{r}_{u,i} - r_{u,i})^2} \qquad \text{(eq:2.5.2)}$$

There is an emerging understanding that good recommendation accuracy alone does not give users of recommender system an effective and satisfying experience, the performance of RSs, must be evaluated with respect to specific user tasks, domain context and also usefulness, that move beyond accuracy to include suitability of recommendations to the user [35]. Suitability includes many terms. Coverage that refers to the proportion of items that RSs can recommend.

Learning rate, measured by the amount of learning data available, some algorithms only needs few data point to start generating acceptable recommendations, while others may need an extensive data point.

On the other end of the spectrum, we find measures like novelty and diversity. Novelty, means recommendations for items that the user did not know about, novelty obviously can be measured by asking users whether they were familiar with a recommended items.

Diversity, is generally defined as the opposite of similarity. In some cases suggesting a set of similar items may not be as useful for the user, because it may take longer to explore the range of items. Furthermore, trust, serendipity, utility, robustness and privacy, can be used to evaluate any recommender system according to the user goal and the domain in which the system is being deployed.

2.6 Human behaviours and Personality

This section dedicated to proving the relationship between human decision and human personality. Personality has been defined as an individual's characteristic pattern of thinking, feeling and behaving, together with the psychological mechanism deriving those patterns [42], thus, makes personality influences how people make their decision. Indeed, it's proved in the literature that people with similar personality characteristics are likely to have similar preferences, [43–47].

The human personality can be modeled by following one of the well known models. The Five Factor

Model (FFM), or the Big-5 personality traits [43], that establishes five broad domains or dimensions called factors, in order to achieve a better description of human personality, these factors can be defined as:

- *Openness (OPE):* reflects a person's tendency to intellectual curiosity, creativity and preference for novelty and variety of experience.
- *Conscientiousness (COS) :* reflects to person's tendency to show self-discipline and aim for personal achievements, and to have an organized and dependable behavior.
- *Extraversion (EXT) :* reflects a person's tendency to seek stimulation in the company of others, showing sociability, talkativeness and assertiveness traits, and to put energy in finding positive emotions, such as happiness, satisfaction and excitation.
- *Agreeableness (AGR) :* reflects a person's tendency to be kind, concerned, truthful and cooperative towards others.
- *Neuroticism (NEU) :* reflects a person's tendency to experience unpleasant emotions, such as anger, anxiety, depressions and vulnerability and refers to the degree of emotion and stability and impulse control.

Measuring the Big-five dimensions (factors) usually done by ”items” that are commonly introduced to the respondent in style of questions. Several types of items has been developed [48], such as the most comprehensive instruments by Costa and McCraeos (1992) [49],240-item NEO Personality Inventory, which permits measurement of the BigFive domains and six specific facets within each dimension, and the wellestablished and widely used instruments,the 44item BigFive Inventory (BFI see [50]; [51], the International Personality item Pool (IPIP) [1] that is publicly available, contains an inventory of all items.

The main reason for using personality in RSs is that, it is a predictable and stable factor, that explains human behaviours, and influences a person's decisions. Gonzalez'z el al [52] emphasize that emotional aspects can be used in the User Profile in order to personalize recommendations in Recommender Systems, and their results showed an improvement in the degree of recommendation acceptance by the user. In addition, Microsoft researcher [53] showed that the five personality traits can play a

[1] International Personality Item Pool, https://ipip.ori.org/newMultipleconstructs.htm

significant role to the improvement in the recommendation quality. In the upcoming chapters, we will prove, and add to the RSs literature that, there is a significant relation between the individual's personality and the choice of the country to live in.

Conclusion

In this chapter, we have introduced, recommender systems notions, defined different recommender systems approaches, followed by the well known methods to evaluate a recommender system algorithms, then we concluded by related works regarding the relation between the person's decisions and their personality. In the next chapter a detailed description of the methodology that we have proposed will be presented.

Chapter 3

Methodology

Introduction

This chapter, presents a full description of the methodology that we have proposed in order to develop and implement a personalized country recommender system. For our experiment, we started by collecting the initial training dataset about countries from real users, in order to feed the collaborative filtering algorithms that we choose. After that we have built a system, that allows users to register, then guides them along a set of steps in order to calculate the recommendations based on each active user rating's, and allow us to collect the needed data to answer our research questions and conduct the analysis phase where we derive patterns and draw conclusions.
The live system is accessible at : http://yurdest.com/
The source code is hosted in a github repository : [54]

3.1 Data Description

In any data based project, collecting a suitable and enough training data is as crucial as developing an accurate model [55]. The scope of initial training dataset intended to be used in this project, was collected using a survey, in which the respondent was asked to fill out, a google form [1] as depicted in Figure 3.1, allowing us to gather enough explicit user ratings about some countries.

[1] https://forms.gle/rsW8QbqZYn8Yy47n7

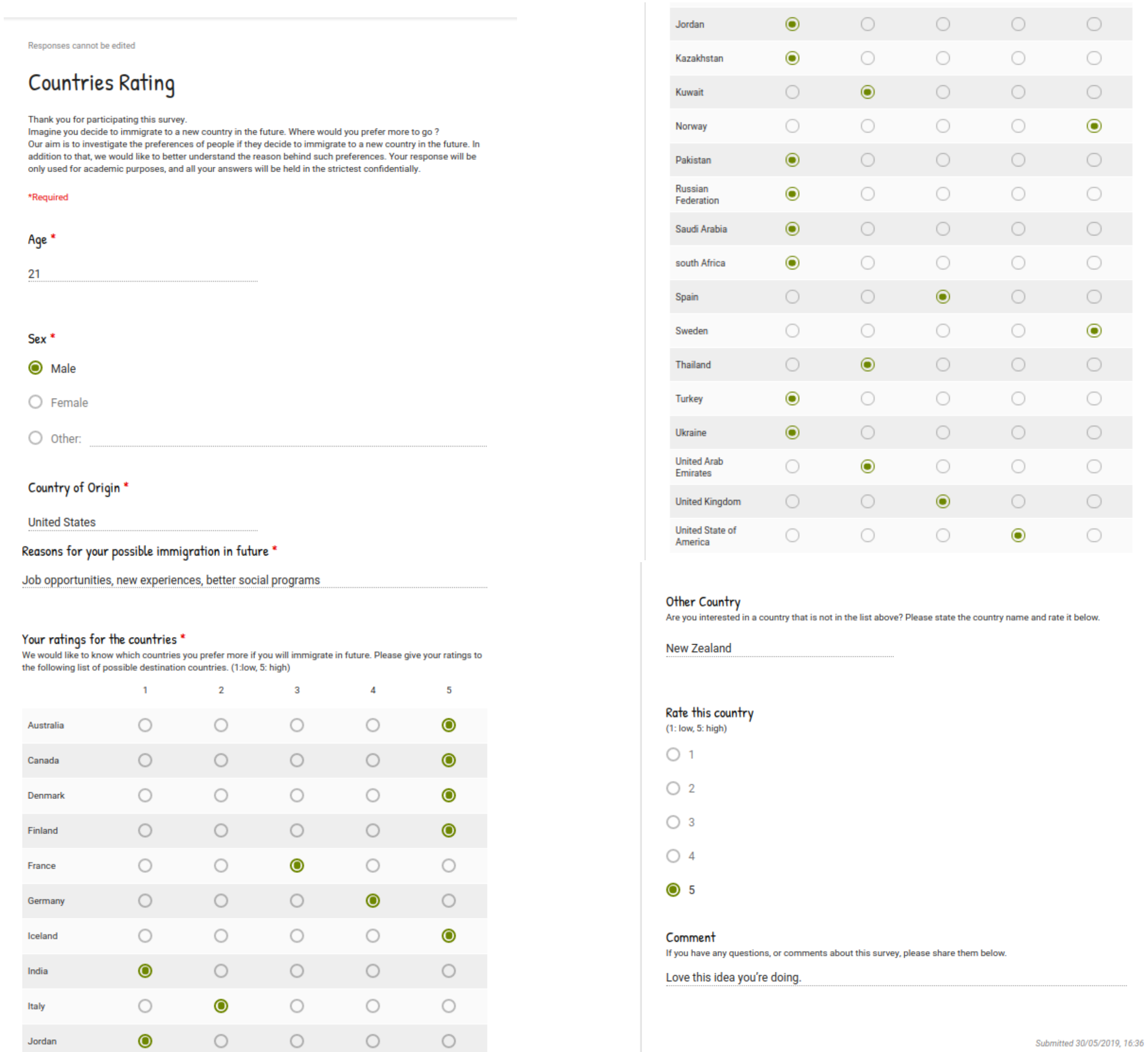


Figure 3.1: Form used in training dataset collection

During 13 days, 136 respondents were taking the survey, each one of them rated at least more than 20 countries using numerical rating (1-5). The selection of countries that was presented to the respondent during the survey in order to rate, was based on the yearly migration report [56] of 2018, by the International Organization for Migration (IOM), that present top destination countries of international migrants.

3.1.1 Data preprocessing

In order to well understand our dataset, some data analysis is required. Firstly the age of the respondent, varied between 14 years old and above 50 years old as shown in Figure 3.2. That could suggest that age can play a significant role for a person to decide where to spend the rest of their life.

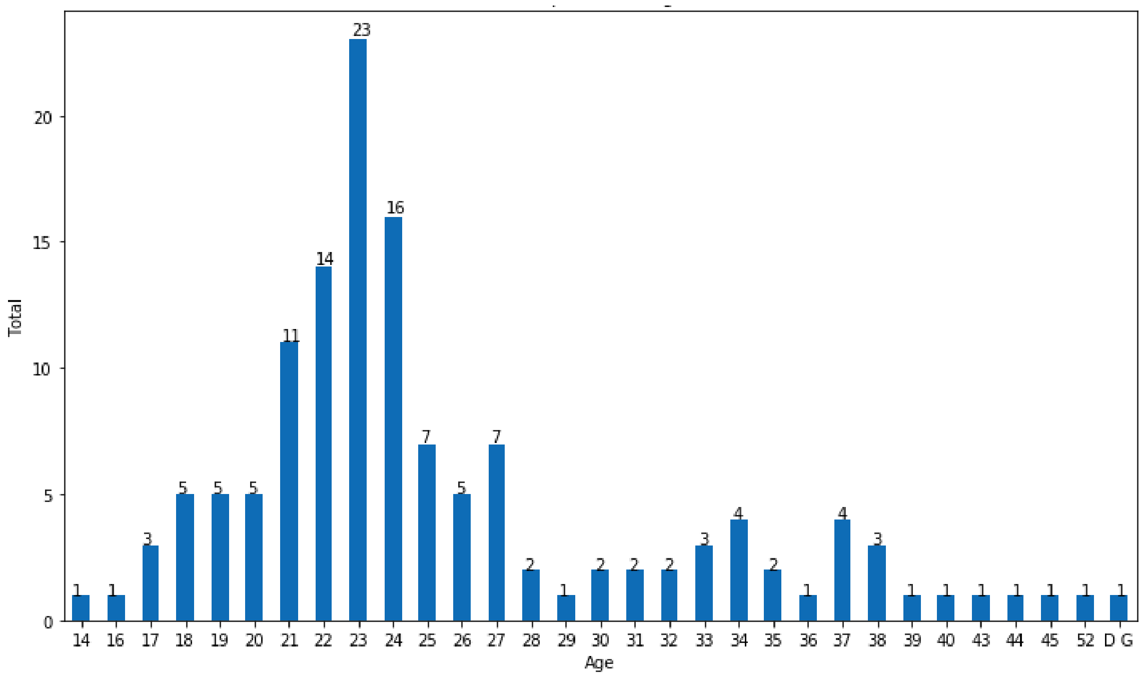


Figure 3.2: Respondents Age

In addition, there is a gender gap, with a higher percentage of male (71.3%) than female (28.%) respondents as illustrated in figure 3.3.

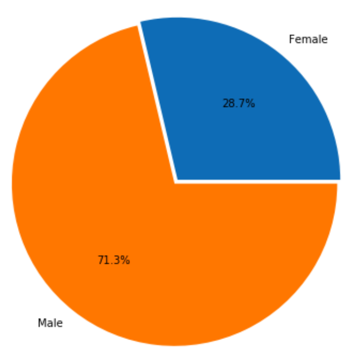


Figure 3.3: Respondents Gender

As expected, almost 47% of the respondent are Moroccan followed by American as illustrated in figure 3.4.

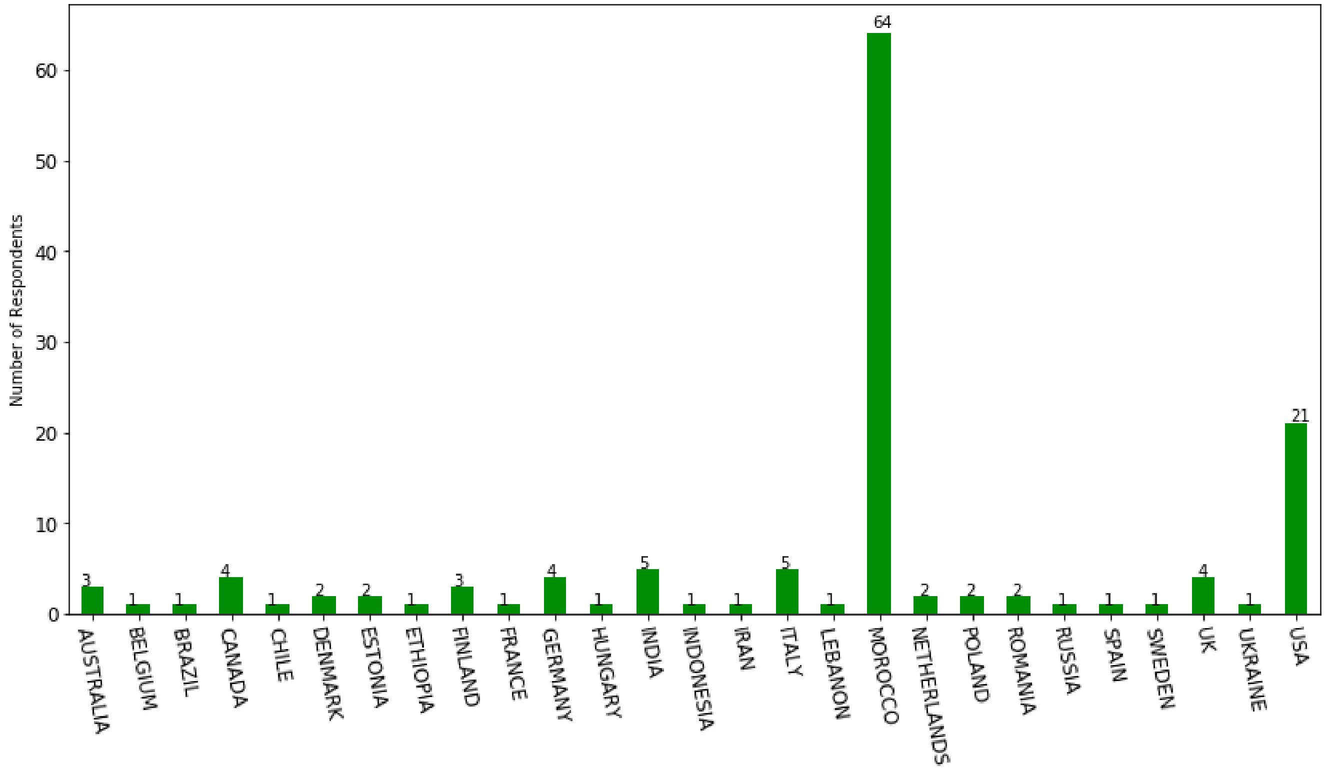


Figure 3.4: Respondents Origin Country

Any one deciding to immigrate to another country, is motivated by a reason, The survey helped us understand why most people decide to emigrate . People emigrate for different reasons, therefore we decide to integrate in our system, a step that allows active users selecting why they seek to move towards new country, as a user preference (3.1). Example of some factors provided during the survey: Job opportunities, new experience, studying or working, find better life condition, seek medical care, political issue, anti-women legislation etc.

Prepossessing the initial training data that will be the input of our algorithms is a crucial step. after the survey was closed, we ended up with 25 selected countries [56] rated explicitly, the table below shows each country and its total rating.

Table 3.1: Countries rating distribution

Country \ Rating scale	5	4	3	2	1
Australia	14	32	38	22	30
Canada	44	38	26	15	13
Denmark	29	33	31	21	21
Finland	29	26	36	23	22
France	13	27	38	37	21
Germany	30	34	35	15	22
Iceland	16	33	36	27	24
India	9	4	20	26	77
Italy	7	24	35	31	29
Jordan	7	4	22	17	86
Kazakhstan	7	4	13	17	95
Kuwait	6	4	18	23	85
Norway	30	37	24	19	26
Pakistan	7	3	13	8	105
Russia	6	10	24	31	65
Saudi Arabia	5	12	29	24	66
Spain	12	21	43	4	20
South Africa	5	17	29	24	66
Sweden	31	35	36	19	15
Thailand	6	15	32	28	55
UAE	14	12	25	24	61
UK	35	33	31	16	21
USA	36	29	28	15	28
Turkey	11	11	26	35	53
Ukraine	6	11	26	29	64

The resulted rating matrix contains 3400 rows, with 136 users rating 25 countries. An example of the comma separated values(csv) ratings matrix used to feed the developed collaborative filtering models, is illustrated in the table bellow (user_id, country_id, ratings, Timestamp: the time of rating)

Table 3.2: Rating matrix

user_id	country_id	ratings	Timestamp
1221	166	5	30-05-2019 16:36:46
1221	187	5	30-05-2019 16:36:46
1543	232	4	30-05-2019 16:41:47
1543	233	2	30-05-2019 16:41:47
1543	239	1	30-05-2019 16:41:47
1888	313	2	30-05-2019 17:46:57
1888	318	2	30-05-2019 17:46:57
1888	322	1	30-05-2019 17:46:57
1888	332	3	30-05-2019 17:46:57
⋮	⋮	⋮	⋮

The python source code, of the data per processing, and the full ratings matrix (CRdata.csv) can be found in our github repository as an ipython notebook [57].

After a literature review, we found that there are a lot of factors that are considered the main engine of the people's' movement around the world wide. In his publication Clare Cummings [58], affirm that is not easy to examine the factors which determines person's decision to emigrate, and comes up with some factors of emigrations such as such : *property, political insecurity*, *the hope of greater opportunities for work*, *family member in a foreign country* etc. The international migration under the microscope paper by Willekens et al [59] present some other factors like: *health care*, *join a partner*, *education quality and amenities*. In addition, Castelli et al [2], come up with a framework to distinguish the factor that act together to inform the final decision to migrate into three categorise, **macro-factors**, **meso-factor** and **micro-factors**, as illustrated in the figure below :

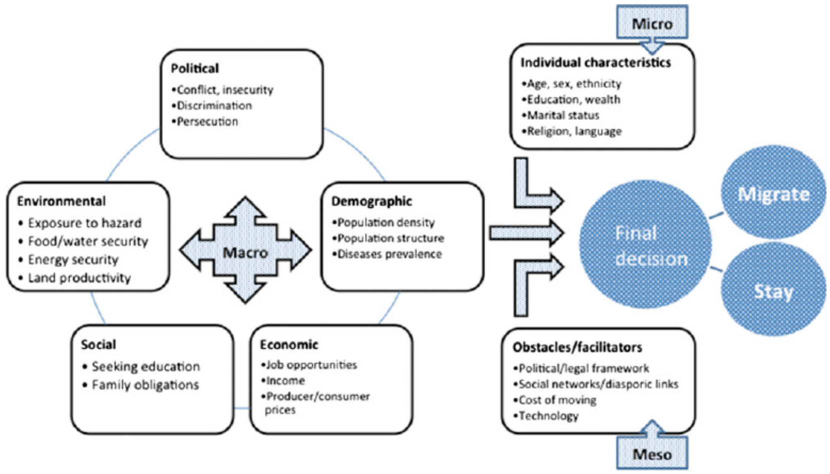


Figure 3.5: Emigration factors [2].

Verwiebe et al [60], study empirically this issue, in order to conduct which factors led German people to move to other country within Europe, by interviewing 40 intermediate qualifications men and women shortly before migration was scheduled in various regions of German, the findings of this study can be listed as : *Age are crucial factor for the decision to immigrate*, *partnership family relations*, *the propose of secure job*, also *living in border regions* and *cultural linguistic similarities and shorter distance* play an important role, more than that extract special immigration motives (European factor) within Europe content such as *information transfer*, *the freedom of movement* and *the mutual recognition of educational certificates and the institutional facilitation of mobility.*

3.2 Recommender Algorithms

One of the important part in this project is figuring out which algorithms would be pertinent in terms of recommendations quality and user satisfaction. Due to the abundance of available RS algorithms, we decided to cross validate [61], several implemented algorithms [62] using Surprise, a Python library for RSs [63]. We relied upon an offline mode. The table below shows the cross-validation results:

Table 3.3: Cross-validation results

Algorithm	RMSE	MAE	Time
Sigulare value Decompusition (SVD)	21.714	16.06	0:00:00
Sigulare value Decompusition++ (SVD ++)	58.217	51.065	0:00:04
Non-negative Matrix Factorization(NMF)	48.659	41.439	0:00:00
SlopeOne algorithm	23.476	18.625	0:00:00
K-Nearest Neighbours (KNN)	21.977	16.88	0:00:00
KNN withMean	21.591	16.493	0:00:00
KNN Baseline	21.589	16.473	0:00:00
Co-clustering	21.393	16.635	0:00:00
Baseline Only	23.505	19.219	0:00:00
Normal Predictor	37.939	30.913	0:00:00

Swiftly we chose three best performing algorithms, according to the lowest combined RMSE (eq:2.5.2) and MAE (eq:2.5.1) measures:

- Singular Value Decomposition (SVD)
- K-Nearest Neighbor Baseline (KNNB)
- Co-Clustring Algorithm

3.2.1 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD), is one of the factorization algorithms for collaborative filtering, and the most successful realization of latent factor models [64] specially after Netflix prize [65].
SVD algorithm, finds the features of users and objects, and makes predictions based on these factors, high correspondence between an item and user factor leads to a recommendation.
Formally, suppose $R \in \mathbb{R}^{n \times m}$ is the ratings matrix, of n user and m items and $I \in 0, 1^{n \times m}$ is its indicator, execution of the SVD algorithm leads to find $M \in \mathbb{R}^{f \times m}$ and $U \in \mathbb{R}^{f \times n}$ as feature matrices of user and objects, with f dimension of feature vector for the user and item. The value of score $R_{i,j}$ is estimated using a prediction function $p(U_i, M_j)$ where U_i and M_j represent the feature vector of user i and object j respectively, the famous prediction function used in matrix factorization is the

dot product , of the feature vectors, $p(U_i, M_j) = U_i^T M_j$, this mean that $R^{m \times n}$ can be factored into:

$$R \approx U^T M$$

The optimization of U and M is performing by minimizing the sum of squared errors between existing scores and their predictions [66].

In this thesis, predicting user ratings $\hat{r}_{ij}$ are based on an implementation of regularized matrix factorization by Simon Funk [67], such that :

$$\hat{r}_{ij} = \sum_{f=1...F} u_{if} m_{jf}$$

where u_{if} denote how much the user likes the factor f and the value m_{jf} denote how strong the factor f is in the item j. Estimation of the factor matrices easily achieved by minimizing the sum of error using regularization and early stopping method techniques [19, 64, 68, 69]:

$$e_{ij} = r_{ij} - \hat{r}_{ij}$$

$$\min \sum_{r_{ij} \in R_{train}} e_{ij}^2 + \lambda(||u_{if}||^2 + ||m_{jf}||^2)$$

and based on straightforward gradient descent:

$$u_{if} \leftarrow u_{if} + \gamma(e_{ij} - \lambda u_{if})$$

$$m_{jf} \leftarrow m_{jf} + \gamma(e_{ij} - \lambda m_{jf})$$

such that, the user and items factor are randomly initialized according to a normal distribution, γ learning rate and λ regularization term.

3.2.2 K-Nearest Neighbor Baseline (KNNB)

K-Nearest Neighbor (KNN) techniques, are amongst the simplest of all machine learning algorithms. As a matter of fact, KNN are in the same context of the idea of CF, that means finding like-minded users (or similar item) is essentially equivalent to finding neighbors for a given user or item [3]. KNN for CF are based usually on three component :

$\rightarrow$ A similarity measures

→ A function that fetch the neighborhood using the similarity measures

→ A rating prediction function based on the neighbor ratings.

In this thesis, based on an implemented KNN Baseline, the ratings $\hat{r}_{u,i}$ for the user u and item i is predicted as follows:

$$\hat{r}_{u,i} = \overline{r}_u + \frac{\sum_{u' \in N_{i(u)}} sim(u,u')(r_{u',i} - \overline{r}_{u'})}{\sum_{u' \in N_{i(u)}} |sim(u,u')|}$$

where $\overline{r}_u$ denote the overall average ratings of user u (baseline estimate) [70], sim(u, u') is a similarity measure between two users u and u', in our case we use cosine similarity (2.4.1.1) measure and $N_{i(u)}$ is a set of users similar to user u who rated the item i.

3.2.3 Co-Clustering Algorithm

Co-Clustering , is a term in data mining that relates to simultaneous clustering of the rows and columns of a matrix. While classical clustering methods assume that a membership of an object depends solely on its similarity to other objects of the same type, co-clustering can be seen as method of co-grouping two types of entities simultaneously, co-clustering a ratings matrix can be explained as grouping both similar user and similar items into, let's say, categories, synchronously [71, 72].
The prediction $\hat{r}_{ui}$ can be computed by assigning the users and items to some clusters C_u, C_i and some co-cluster C_{ui}:

$$\hat{r}_{ui} = \hat{C}_{ui} + (\mu_u - \hat{C}_u) + ((\mu_i - \hat{C}_i))$$

where $\hat{C}_{ui}$ is the average rating of co-cluster C_{ui}, $\hat{C}_u$ is the average rating of u's cluster, and $\hat{C}_i$ is the average ratings of i's cluster, and Clusters are assigned using a straightforward optimization method.

3.3 Implementation and Design

3.3.1 Implementation

We have deployed a dynamic online system for collection of the user data and providing recommendations. An example of such system is proposed in [73]. We use three collaborative filtering algorithms

(3.2) to personalise recommendations for the active user. Recommendations are computed, in real time after the user submits his ratings of a few countries that is interested in. The main goal of the system is to implement the existing algorithms and gather enough data for further analysis, and not implementing or coming with new design.

The system was built using the well known open sourced, full-stack Python framework, Django version 2.2 [74] and a custom JS/CSS/HTML for the front end implementation, a personal computer with 2.50 GHz Intel(R) Core(TM) i7-6500U CPU processor and 8GB of RAM. The whole system is deployed with apache2 installed in Unix Server with 1GB of RAM, 2GH Intel Xeon E3-12xx v2 (Ivy Bridge, IBRS), and a PostgerSQL database management system [75] to hold user data, countries data, ratings data, survey data and final recommendations.

3.4 Architecture Decomposition

The user interacts with the system, via web application. There is a sequence of steps/web pages, each represent a form that the user should fill out and submit, in order to move to the next steps and save the data. The ratings step, give the users freedom to rate any country they want without any restrictions, after features selection (3.5.1.3), in order to communicate the ratings with the algorithms and get three lists of recommendations, each corresponding to a separate algorithm result: SVD, KNN Baseline, Co-clustering (3.2), then the recommendation lists are saved into the database, which makes it easier for the user to fetch them in the next login. The system architecture is illustrated in the figure below.

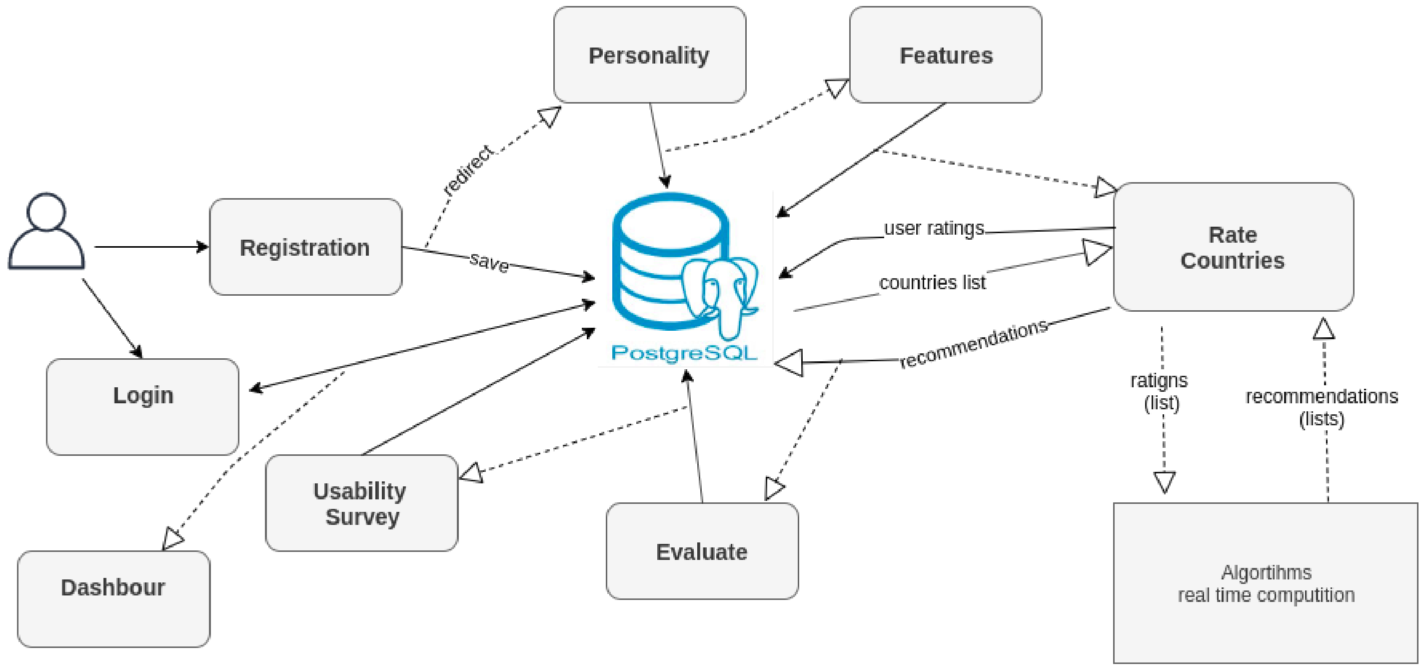


Figure 3.6: System Architecture

3.5 Design

The user interface design is the design of the graphical elements on the computer screen with which a user interacts to conduct application tasks. The user interface accomplishes two fundamental tasks: communicating information from the computer to the user and from the user to the computer. We have tried to make the user experience as smooth and straightforwards as possible through effective user interface.

3.5.1 User Flow

User flow is the path a user follows through your web application interface to complete a task.

3.5.1.1 Authentication

In the initial step we provide the user with some information about our project, and we ask them to register to the system in order to make it more personalized, providing us with some information:

- username
- email
- password

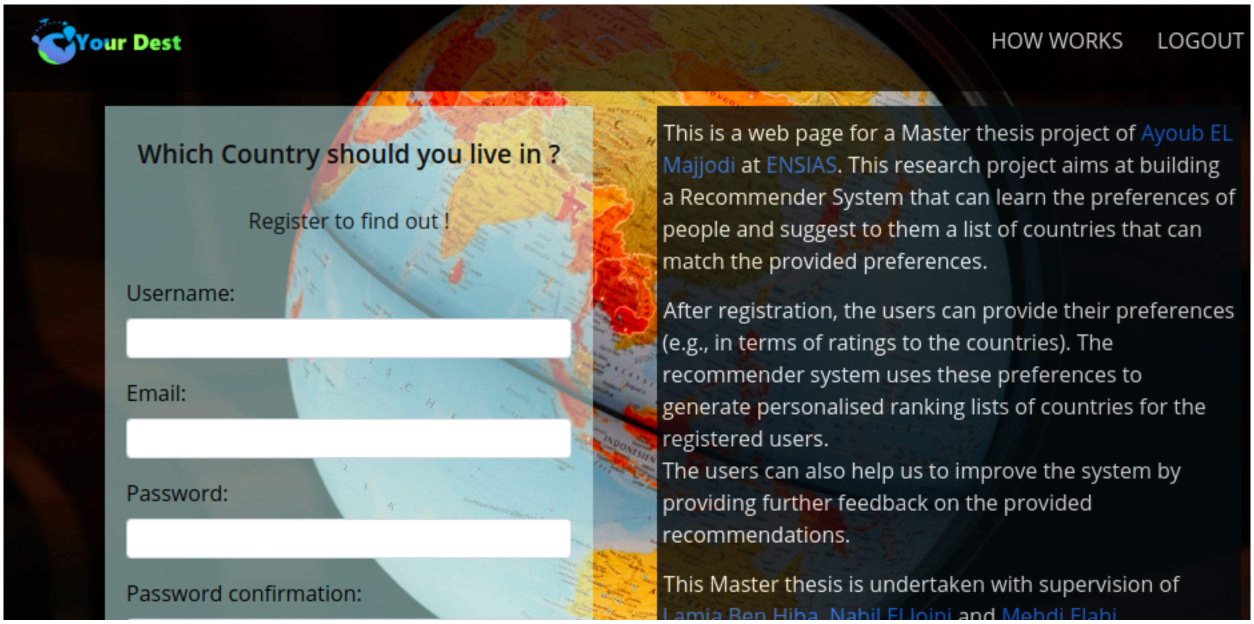


Figure 3.7: Authentication

3.5.1.2 Personality questionnaire

After user registration, we present them with a personality questionnaire, age, gender, country of origin, and five questions related to the Big-5 personality traits (2.6),Openness(imaginative), conscientiousness (organized), extraversion (enthusiastic), agreeableness (kind), neuroticism (kalm).

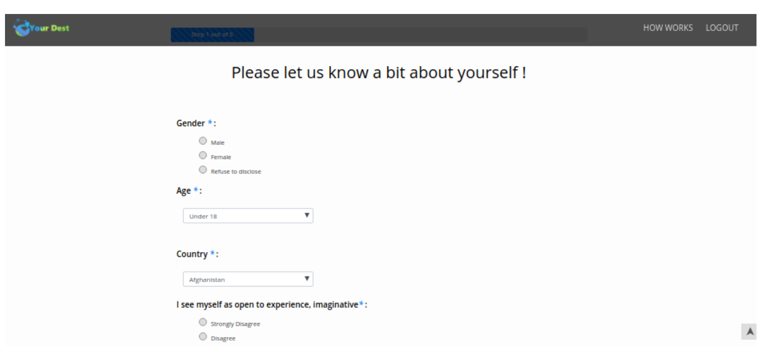


Figure 3.8: Personality questionnaire

3.5.1.3 Feature Selection

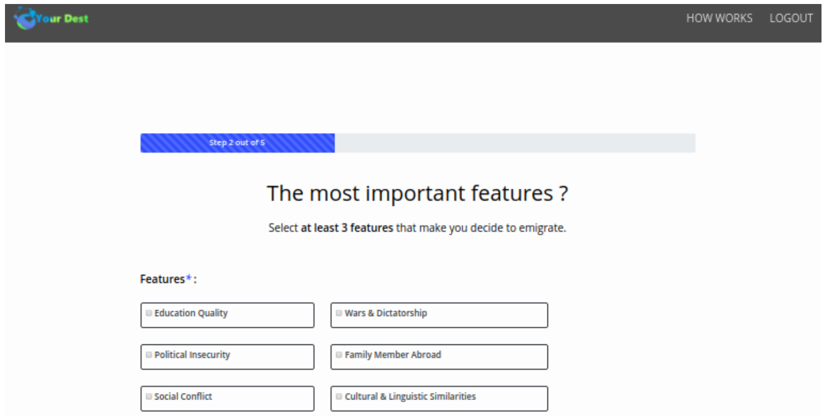


Figure 3.9: Features selection

Features selection step, considered one of the most important steps of the social side of our project, we impose that the users choose at least three factors that make them decide to emigrate.
In this thesis, we present to the user the most cited factors in the literature and the factors that have an open dataset. The dataset about each factor can be found in [76]. We ended up with 12 factors :

- Education quality
- Political insecurity
- Social conflict
- Work opportunities
- Health care
- Income difference
- Wars and dictatorship
- Family member abroad
- Cultural and linguistic similarities
- Working atmosphere
- Shorter distance
- Crime rate

3.5.1.4 Ratings

Next we perform rating elicitation [77, 78] by requesting the user to select a number of countries and rating them. If the user prefers to rate more countries, we present them with "ADD MORE" button to add more countries and ratings to the chart, The user should rate at least 5 countries using a rating scale from 1 to 5.

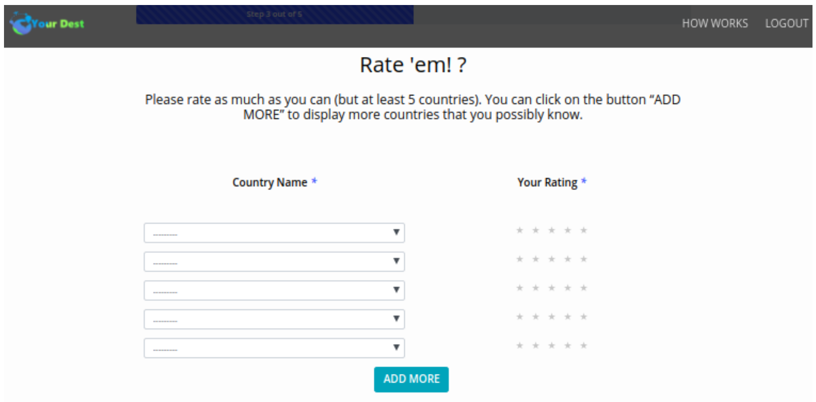


Figure 3.10: Ratings step

3.5.1.5 Results

After the ratings are submitted, we present three lists of personal country recommendations, each list displays the recommendation algorithm we employed (SVD, KNN, Co-clustering). We also present 14-question survey to evaluate the lists, with regards to their usefulness and recommendations quality (2.5).

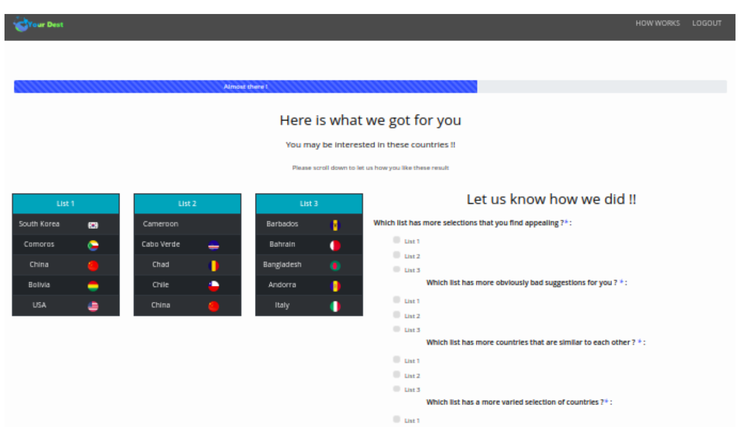


Figure 3.11: Results Evaluation

3.5.1.6 Final step

In this step, we present 10 question survey to the users in order to let them evaluate the system according to The system Usability Scale (SUS) [79].

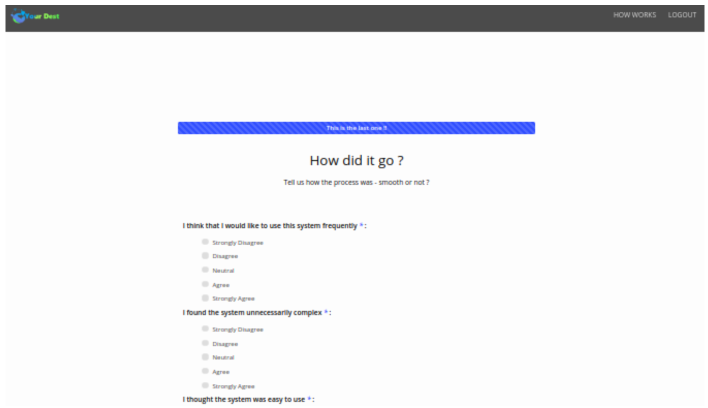


Figure 3.12: Usability survey

conclusion

This chapter started by the presentation of the training dataset in order to get insight and prepare the data that are collected by a question survey, after that, we gave the theoretical foundation of the recommendation algorithms, that we are implemented in this project, at the end we illustrated the design system and implementing details. The next chapter will be dedicated to the experiment and the results analysis.

Chapter 4

Experiments and Result

Introduction

In this chapter we will provide the details of the set of experiments that we conducted in order to investigate the effectiveness of our system, then we will discuss the main results of this research work.

4.1 Experiments

On-line experiments on real users can help evaluate the quality of any personalized recommender system and allow direct measurement of overall system goals [41, 80]. Prior the experiments, a small dataset of a few hundred of user preferences in terms of ratings was available from a pre-survey (3.1), the dataset was about 3400 ratings for 25 countries, that has been imported into the database and used to bootstrap the recommendation algorithms. Before preforming our experiments, the entire system components have been designed, developed and deployed. After initial tests, we shared the link to reach a wide range of users. The participants registered in the system, provided their preferences, evaluated the generated recommendations and provided feedback on their experience during the interaction with the system regarding the system usability.

In general, we received ratings from 193 user. The mean was 7 ratings per user, while the median was about 6 ratings, the minimum was 5 ratings and the maximum was 41 ratings, 1284 was the total number of ratings.

4.1.1 Users

The system received 281 registered people, but only 109 completed all the steps. Unfortunately we lost a lot of respondents after the registration step (41 respondents), some respondents did not want to share their personal information. We also lost a sizeable amount of people after the lists evaluation step (80 people). Obviously some users did not go through the system evaluation step, either due to the lack of time or due to being discouraged by the task. Table 4.1 illustrate how many users attempted each step of the experiment.

Table 4.1: User attempted the survey

Registered Users	281 (100%)
Users Completed Personality Survey	241 (85%)
Users Completed Features Survey	226 (80%)
Users Submitted Rating	193 (69%)
Users Completed Evaluation Survey	189 (67%)
Users Completed Usability Survey	109 (38)%

Among our users, the male gender prevailed with 170 respondents, 65 female and 5 people refused to disclose their gender. Table below shows the respondents gender.

Table 4.2: User demographics: gender

Males	**Females**	**Refused to disclose**
170 (71%)	65 (27%)	5 (2%)

The majority of our respondents were between, the age of 18 and 24, almost 42% of all the registered users. Table 4.3 show the age distribution of the respondents.

Table 4.3: User demographics: age

Under 18	**18-24**	**25-35**	**35-45**	**45-55**	**Over 55**
12 (5%)	100 (42%)	93 (39%)	23 (9%)	10 (4%)	2 (1%)

Majority of our respondents were from USA, next larger group were from Morocco and Egypt. Table 4.4 show the users' origin country distribution.

Table 4.4: User demographics: country of origin

Number of Respondents per Country	**Country of Origin**
56	USA
32	Morocco
15	Egypt
10	Canada, UK
8	Germany
7	Australia, Italy, Kuwait
6	UAE
5	India
4	Turkey, France, Afghanistan
3	Brazil, Mexico, Spain, Netherlands, Belgium Qatar, Sweden, Hungary
2	Ireland, Norway, Ukraine, Russia, Singapore Czech Republic, Pakistan, Indonesia
1	Japan, Sri Lanka, Austria, Nigeria Iran, Dominica, Tunisia, Korea, Croatia Switzerland, Lithuania, Chile, Malta Romania,Guiana, Palestine, New Zealand Slovenia, Slovakia, Saudi Arabia, Panama Jordan, Armenia, Portugal, Syrian Arab Republic

Since we shared out a web link, we had no way to control users and their progress, which resulted in the percentage of turnout being less than expected. However, since the users had full autonomy during the survey sequences, we believe it gave them more freedom during the evaluation of the system.

4.2 Results

In this section we discuss obtained results after the experiments. As we mentioned, we had a total of 281 (100%) participants registered, 240 completed the personality survey, 225 took the features questionnaire, 193 submitted the ratings and 189 evaluated the recommendation lists, finally only 38% of the participants completed all the steps.

4.2.1 Ranking List evaluation

We implemented three algorithms that can generate personalized ranking lists of countries. These algorithms were selected relying upon an offline cross-validation study, and top 3 performing have been selected for the real user study.

The personalized ranking lists were generated by the following algorithms: Co-clustering algorithm, k-Nearest Neighbor Baseline algorithm and Singular Value Decomposition algorithm.

We asked the user to fill out a 14 question evaluation survey in order to compare the generated ranking list by these algorithms. Each of the questions evaluate the quality of the ranking lists in terms of 5 different dimensions, i.e., *Accuracy, Diversity, Satisfaction, Novelty* and *Understands Me*. The questionnaire was proposed and validated in a the literature [81]. In fact the original questionnaire was prepared for movie domain and consisted of 22 questions that aimed at evaluating a recommendation list of movie suggestions. For our study, the questions were rephrased and adapted the nature of our propositions. Table 4.5 shows the overview of the results we obtained from the users, about comparing the suggestions list.

Table 4.5: Comparison of the algorithms, generating different country ranking lists

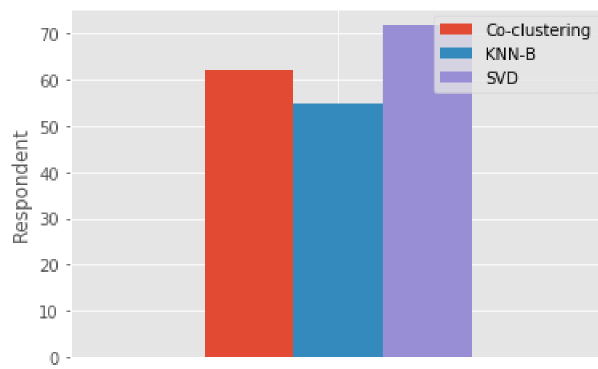
Metric	**Question**	**Co-clustering**	**KNN-B**	**SVD**
Accuracy	Which list has more selections that you find appealing ?	33%	29%	38%
Accuracy	Which list has more obviously bad suggestions for you ?	59%	31%	10%
Diversity	Which list has more countries that are similar to each other ?	26%	26%	48%
Diversity	Which list has a more varied selection of countries ?	42%	40%	18%
Diversity	Which list has countries that match a wider variety of preferences ?	24%	47%	29%
Understand Me	Which list better reflects your preferences in countries ?	18%	26%	56%
Understand Me	Which list seems more personalized to your countries ratings ?	21%	24%	55%
Understand Me	Which list represents more mainstream ratings instead of your own ?	15%	24%	61%
Satisfaction	Which list would better help you find countries to consider ?	14%	40%	46%
Satisfaction	Which list would you be more likely to recommend to your friends ?	19%	19%	62.4%
Novelty	Which list has more countries you did not expect ?	55%	33%	12%
Novelty	Which list has more countries that are familiar to you ?	23%	29%	48%
Novelty	Which list has more pleasantly surprising countries ?	25%	49%	26%
Novelty	Which list provides fewer new suggestions ?	29%	17%	54%

Observing the results, we have noticed that raking list generated by the SVD algorithm, achieves better results compared to other algorithms. Next, we checked the evaluation metrics.

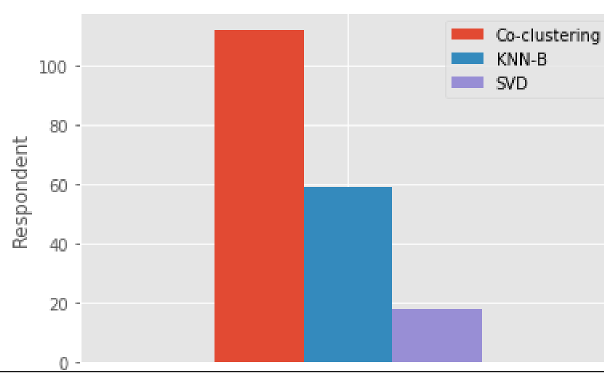
In terms of Accuracy, the users have evaluated higher the ranking list provided by the SVD algorithm, followed by KNN-B and lastly Co-clustering. Table 4.6 illustrates the details of the Accuracy metric result.

Table 4.6: Evaluation survey: Accuracy

<table>
<tr><th>Accuracy</th></tr>
<tr><td>1. Which list has more selections that you find appealing?



</td></tr>
<tr><td>2. Which list has more obviously bad suggestions for you?



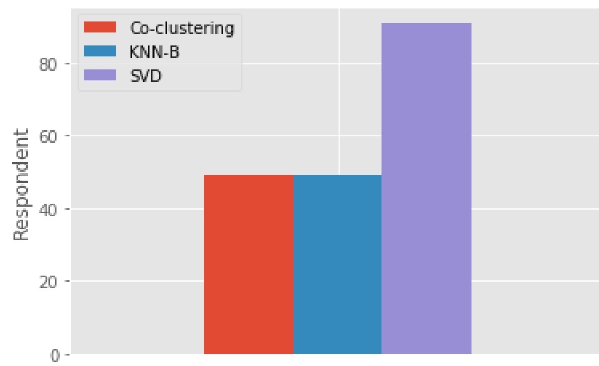
</td></tr>
</table>

In terms of Diversity metric, the users found the countries within the list generated by the SVD algorithm to be more similar to each other. The ranking list generated by Co-clustering was chosen as the list that provide varied selection of countries and for the list that match a wider variety of preferences was the one generated by KNN-B. Table below illustrates the details of the diversity evaluation results.

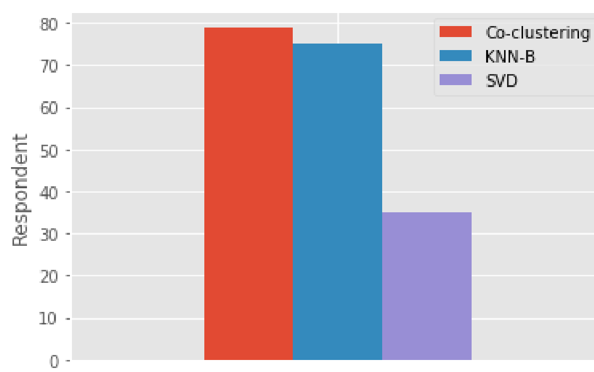
Table 4.7: Evaluation survey: Diversity

Diversity

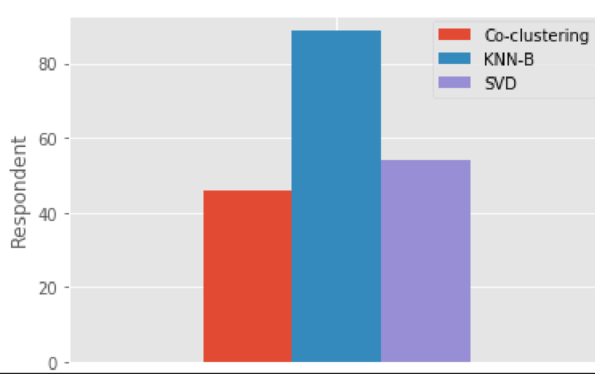
3. Which list has more countries that are similar to each other?



4. Which list has a more varied selection of countries?

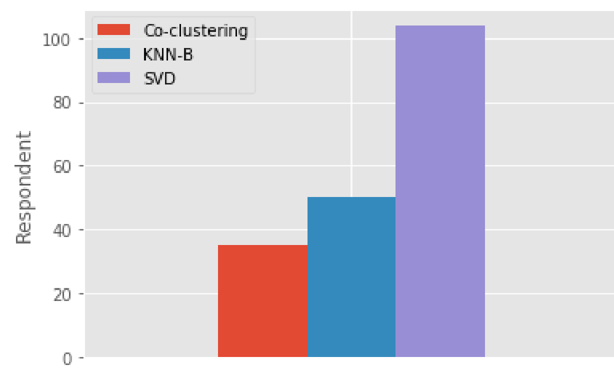
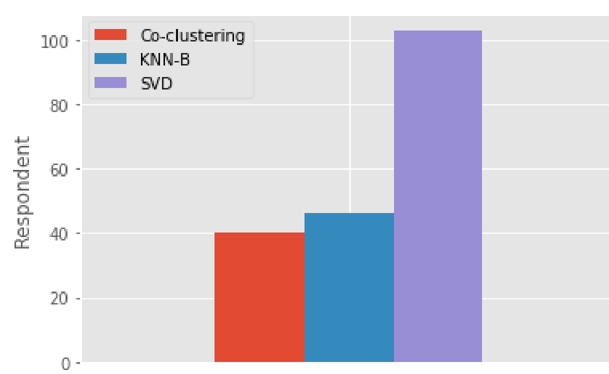
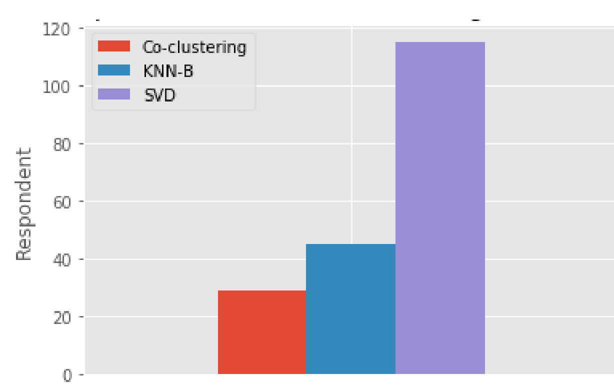


5. Which list has countries that match a wider variety of preferences?



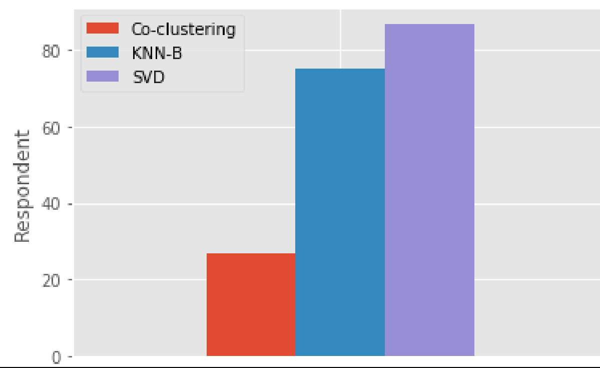
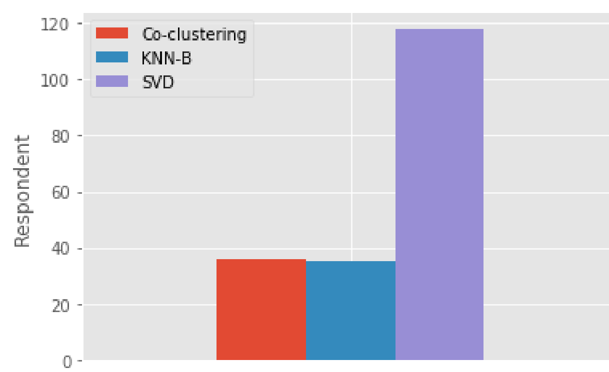
The Understand Me metric, evaluates perceived personalization. The countries ranking list generated by SVD was much more preferred than the other algorithms. This could be due to the fact that SVD may select more mainstream countries than other algorithms, this could be explained by the fact that, top destination countries get higher ratings by many users. KNN-B and Co-clustering comes close to each other, in order to well personalize countries to the users. Table 4.8 illustrates the understand Me factor metric results.

Table 4.8: Evaluation survey: Understand Me

<table>
<tr><th>Understand Me</th></tr>
<tr><td>6- Which list better reflects your preferences in countries ?

</td></tr>
<tr><td>7- Which list seems more personalized to your countries ratings ?

</td></tr>
<tr><td>8- Which list represents more mainstream ratings instead of your own ?

</td></tr>
</table>

For the satisfaction, we see that the user were more satisfied with the result presented by the SVD algorithm and less satisfied with by the Co-clustering result. KNN-B and Co-clustering comes close to each other in terms of predicting the results that can a user recommends to a friend. Table 4.9 shows the satisfaction metric evaluation.

Table 4.9: Evaluation survey: Satisfaction

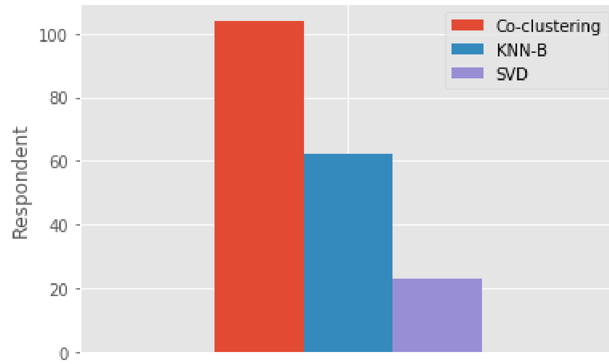
Satisfaction
9- Which list would better help you find countries to consider ?

10- Which list would you be more likely to recommend to your friends ?


As for the Novelty, SVD generated the expected countries, and fewest new recommendations, Co-clustering produced most amount of unexpected countries and the users chose KNN-B algorithm the one that produced more pleasantly surprising countries. Table 4.10 illustrates the result about the Novelty metric.

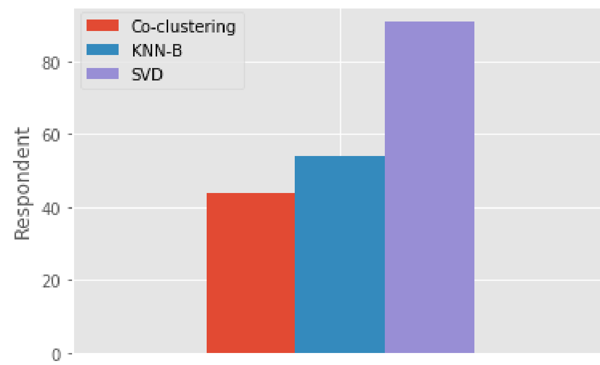
Table 4.10: Evaluation survey: Novelty

Novelty

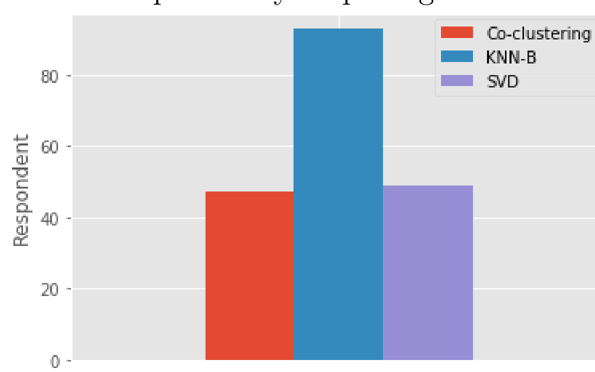
11- Which list has more countries you did not expect ?



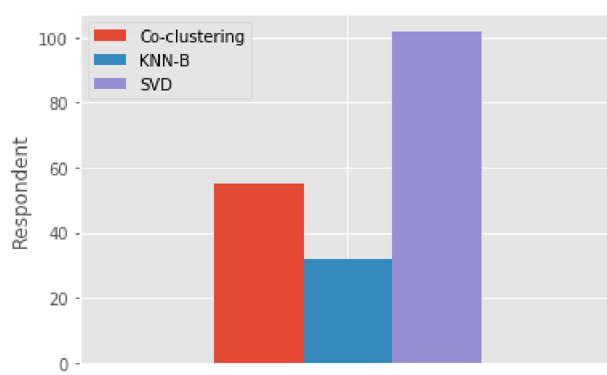
12- Which list has more countries that are familiar to you ?



13- Which list has more pleasantly surprising countries ?



14- Which list provides fewer new suggestions ?



In order to determine if there is a significant difference between the algorithms, we performed a T-test and calculated the p-values for the data. We computed a set of overall p-values, comparing SVD with KNN-B, SVD with Co-clustering and Co-clustering with KNN. Table 4.11 below shows the results received from ruining a paired T-test.

Table 4.11: Algorithm evaluation, (t-test, p-value) $H_0 = 0.05$

Comparison condition	Paired T-test, p-value
SVD vs **KNN-B**	0.10
SVD vs **Co-clustering**	0.076
KNN-B vs **Co-clustering**	0.65

As a conclusion, from the p-values calculated we can see that there is no significant difference between SVD and KNN-B, and there is a slight significant between SVD and Co-clustering, but also no significant difference between KNN-B and Co-clustering.

4.2.2 Personality Analysis and Features Preferences

Next we investigated how the Personality Survey responses and the Features Survey responses can be correlated. From the 240 people who completed the personality questionnaire, only 255 took the feature survey. The features used were taken from the literature (cf. 3.1.1).

Firstly, we viewed that the important feature among the people was **Work opportunities**, that was chosen by almost 72% of the respondent, followed by **Education quality** and **Working atmosphere** with 46% of the votes and 44% of the votes respectively. The **Health care** feature comes third with 43% of the total choice by respondents. Participants were most concerned with the features that most correlated with the quality of life. The least chosen feature was **Family member abroad** and **Shorter distance** with 8% and 6% of the votes respectively which was previously emphasized by [60]. Table 4.14 illustrates the votes about the features.

Table 4.12: Votes of Features Preferences

Features	**Respondent choose it**
Work Opportunities	161 (72%)
Education Quality	105 (46%)
Working Atmosphere	100 (44%)
Health Care	97 (43%)
Income Difference	84 (37%)
Political Insecurity	59 (26%)
Crime Rate	58 (26 %)
Social Conflict	49 (22%)
Cultural & Linguistic Similarities	47 (21%)
Wars & Dictatorship	37 (16%)
Family Member Abroad	20 (8%)
Shorter Distance	15 (6%)

Next we looked at the correlation between the features and the gender of the respondents. We saw that **Working opportunities** feature is the most considered one by people when choose a country to live in. In addition, males tended to select **Education quality** more often than female, 48% of males choose this feature versus 33% of females. Also male respondents care more about **Income differences** feature than female. Our female participants were more concerned with **Family member abroad** than male participants.

Table 4.13 shows which features were selected by those of our participants who considered themselves as females.

Table 4.13: Feature Preferences by Gender: Females

Features	**Respondent choose it**
Work Opportunities	40 (61%)
Working Atmosphere	31 (48%)
Health Care	24 (37%)
Education Quality	22 (33%)
Political Insecurity	14 (21%)
Crime Rate, Social Conflict, Family Member Abroad	12 (18%)
Income Difference	11 (18%)
Cultural & Linguistic Similarities, Wars & Dictatorship	9 (14%)
Shorter Distance	4 (6%)

Table 4.14 shows which features were selected by male participants.

Table 4.14: Feature Preferences by Gender: males

Features	**Respondent choose it**
Work Opportunities	118 (70%)
Education Quality	81 (48%)
Health Care	72 (42%)
Working Atmosphere	68 (40%)
Income Difference	70 (41%)
Crime Rate	44 (26%)
Political Insecurity	41 (24%)
Cultural & Linguistic Similarities, Social Conflict	36 (21%)
Wars & Dictatorship	26 (15%)
Shorter Distance	10 (6%)
Family Member Abroad	7 (4%)

Next, we split the respondents collected from the personality survey, into two groups, A and B for each question, based on the median of the participants. Group A contains the respondents with the answers higher than the median, which for all 5 personality questions were those who choose *Agree* or *Strongly Agree* option for a given single question, while Group B included the respondents with answers lower than the median, those who chose *Disagree*, *Strongly Disagree* or *Neutral*. The table below shows the groups of participants.

Table 4.15: A/B Groups split

Personality question	**Group A**	**Group B**
1. **Openness:** I see myself as open to experience, imaginative.	Strongly Agree Agree 83%	Strongly Disagree Neutral Disagree 17%
2. **Conscientiousness:** I see myself as dependable, organized	Strongly Agree Agree 68%	Strongly Disagree Neutral Disagree 30%
3. **Extroversion:** I see myself as extroverted, enthusiastic.	Strongly Agree Agree 57%	Strongly Disagree Neutral Disagree 43%
4. **Agreeableness:** I see myself as agreeable, kind.	Strongly Agree Agree 75%	Strongly Disagree Neutral Disagree 25%
5. **Neuroticism:** I see myself as emotionally stable, calm.	Strongly Agree Agree 62%	Strongly Disagree Neutral Disagree 38%

After splitting the respondents into two groups, we looked at the demographics and what the top 3 most selected features for each group are.

As we see from table 4.16, the group of respondents, those who consider themselves very open to experience and imaginative, were leaning towards *Work opportunities*, *Health care* and *Education quality*, for those who identified themselves as not imaginative gave much more votes to *Health care* and *Income differences* than *Education quality.*

Table 4.16: **Question 1. I see my self as open to experience, imaginative**

Category	**Group A**	**Group B**
Answers	Strongly Agree Agree	Strongly Disagree Neutral Disagree
Respondents	190	36
Top selected features	Work opportunities(70%) Education quality(47%) Health care(41%)	Health care(47%) Income Difference(47%) Education quality(41%)
Age	18-24	18-24
Gender	Females (27%) Males(72%)	Females (20%) Males(72%) Refused to disclose(8%)

From the Table 4.17, both groups of participants, those who consider themselves very dependable and organized and those who do not, cared much more about *Working opportunities*, *Working atmosphere* and *Education quality* with slightly different votes.

Table 4.17: **Question 2. I see myself as dependable, organized**

Category	**Group A**	**Group B**
Answers	Strongly Agree Agree	Strongly Disagree Neutral Disagree
Respondents	142	87
Top selected features	Work opportunities (71%) Working atmosphere (45%) Education quality (45%)	Work opportunities (68%) Working atmosphere (62%) Education quality(47%)
Age	18-24 (40%) 25-35 (35%)	18-24 (44%) 25-35 (32%)
Gender	Females (25%) Males(70%)	Females (26%) Males(71%)

Those who consider themselves more extroverted and enthusiastic, firstly, were interested in *Working opportunities* with 69% of the votes, versus *Education quality* and *Working atmosphere* with 53% and 48% of the votes respectively. Those who are not enthusiastic, gave many more vote to *Working opportunities* (73%) than *working atmosphere* and *Education quality* both with 42% of the votes. the Table bellow depicts more details about enthusiastic groups.

Table 4.18: **Question 3. I see myself as extroverted, enthusiastic.**

Category	**Group A**	**Group B**
Answers	Strongly Agree Agree	Strongly Disagree Neutral Disagree
Respondents	90	136
Top selected features	Work opportunities (69%) Education quality(53%) Working atmosphere (48%)	Work opportunities (73%) Working atmosphere (42%) Education quality (42%)
Age	18-24 (41%) 25-35 (39%)	18-24 (43%) 25-35 (31%)
Gender	Females (23%) Males(74%)	Females (27%) Males(70%)

From Table 4.19, both agreeable and less agreeable groups were interested in *Work opportunities* and *Education quality*. Group A pays much more attention to *Education quality*, while group B is more concerned with *Health care*.

Table 4.19: **Question 4. I see myself as agreeable, kind.**

Category	**Group A**	**Group B**
Answers	Strongly Agree Agree	Strongly Disagree Neutral Disagree
Respondents	170	56
Top selected features	Work opportunities(69%) Working atmosphere (47%) Education quality(45%)	Work opportunities(77%) Education quality(56%) Health care(43%)
Age	18-24 (43%) 25-35 (36%)	18-24 (41%) 25-35 (38%)
Gender	Females (27%) Males(71%)	Females (23%) Males(73%)

Regarding emotional stability and calmness, both groups were interested in *Work opportunities* and *Education quality*, but group A was leaning towards *Health care* and group B towards *Working atmosphere*. Table 4.20 illustrates the details about those groups.

Table 4.20: **Question 5. I see myself as emotionally stable, calm.**

Category	**Group A**	**Group B**
Answers	Strongly Agree Agree	Strongly Disagree Neutral Disagree
Respondents	142	84
Top selected features	Work opportunities(73%) Health care(44%) Education quality(48%)	Work opportunities(69%) Working atmosphere (50%) Education quality(44%)
Age	18-24 (46%) 25-35 (35%)	18-24 (36%) 25-35 (39%)
Gender	Females (20%) Males(78%)	Females (36%) Males(60%)

4.2.3 The Personality and Algorithms Preferences

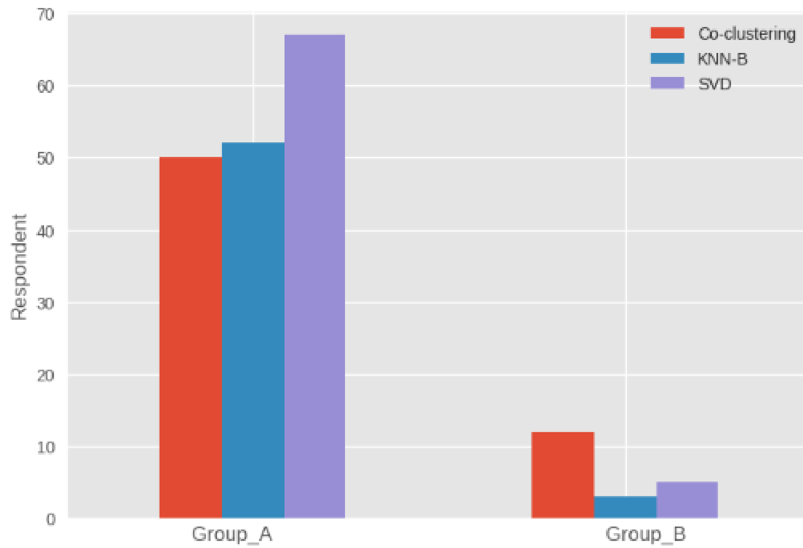
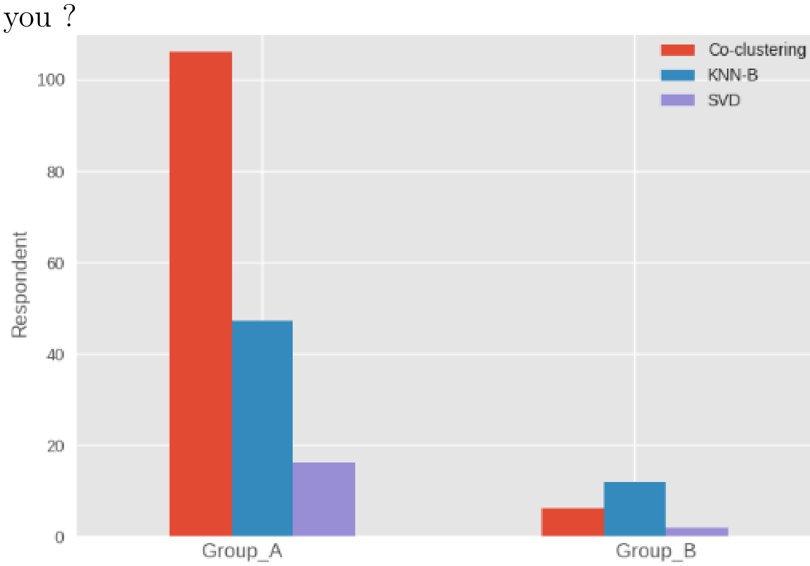
In this section we will, investigate the correlation between the personality data we collected and the responses to the Evaluation Survey, we asked the users to evaluate the generated list of recommendations according to the five metrics : *Accuracy, Diversity, Understand Me, Satisfaction and Novelty.*

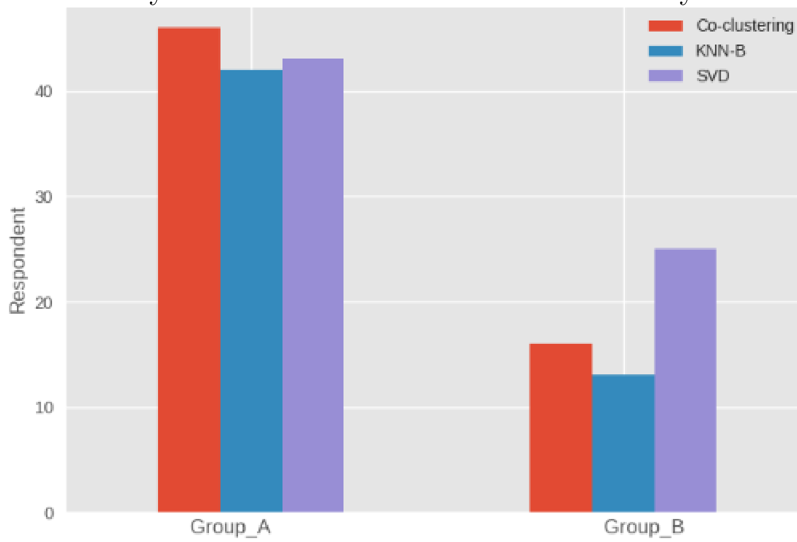
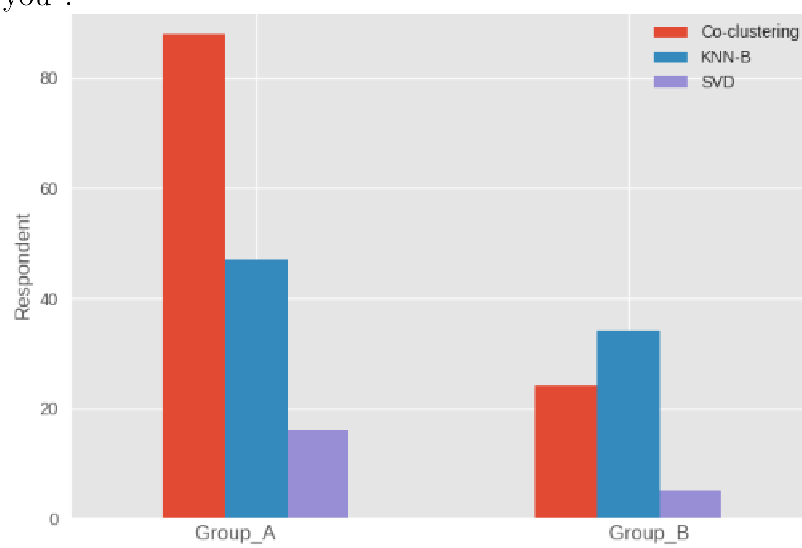
As we mentioned, 241 respondents completed the Personality Survey, while only 109 completed the Evaluation Survey, we focused only on the participants who completed both surveys, and per each personality question we split the participant into two groups A and B, just as we did in the previous section. Table 5.18 in the previous section shows the details of the group split. Below are the results we obtained grouped by list Evaluation Survey metrics.

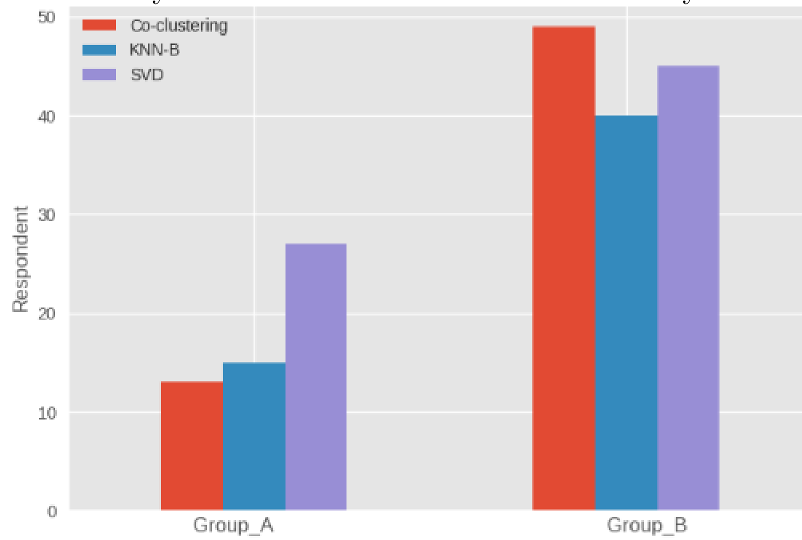
4.2.3.1 Accuracy Metric

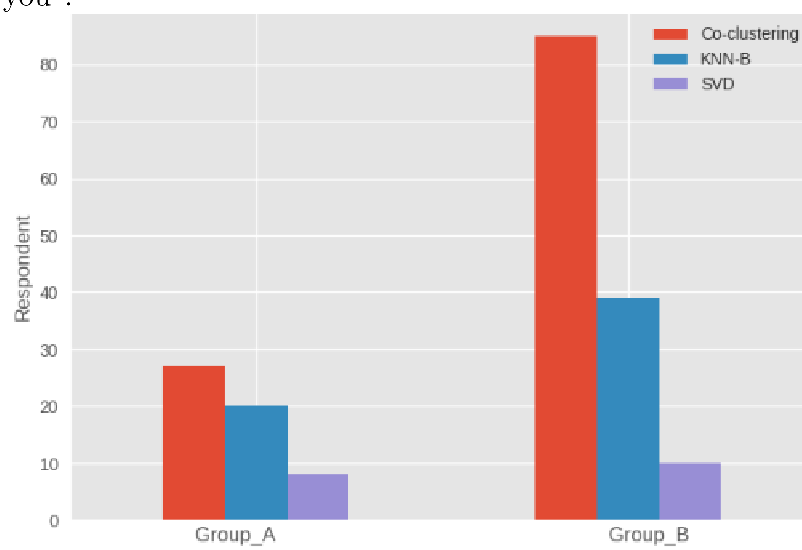
Table 4.21 shows the difference in users' assessment of the algorithms accuracy. We can see that groups of participants, prefer the SVD algorithm in terms of accuracy. For the group A (highly imaginative and open to experience) the KNN-B algorithm comes second in terms of finding appealing selections. Many more members of group A consider the Co-clustering algorithm worse than KNN-B in term of finding good suggestions, for the group B (not open to experience) KNN-B produces by far more obviously bad suggestions.
As for respondents who consider themselves more dependable and organized, they chose co-clustering over SVD in terms of accuracy of suggestions while their counter parts, those who not consider themselves highly organized, clearly they chose SVD.
For those people who consider themselves, extroverted and enthusiastic they chose SVD in terms of accuracy while the group B preferred Co-clustering as most algorithm generated appealing selections and SVD as the most algorithm suggest good recommendations.
Those participants who scored high in agreeableness chose SVD over Co-clustering and KNN-B, in terms of accuracy. For those who do not consider themselves kind, they chose Co-clustering generated appealing selections and more obviously bad suggestions.
Both participants who consider themselves emotionally stable and calm, and the participants who do not consider themselves as such preferred SVD over KNN-B and Co-clustering in terms of accuracy.

Table 4.21: Accuracy and Personality

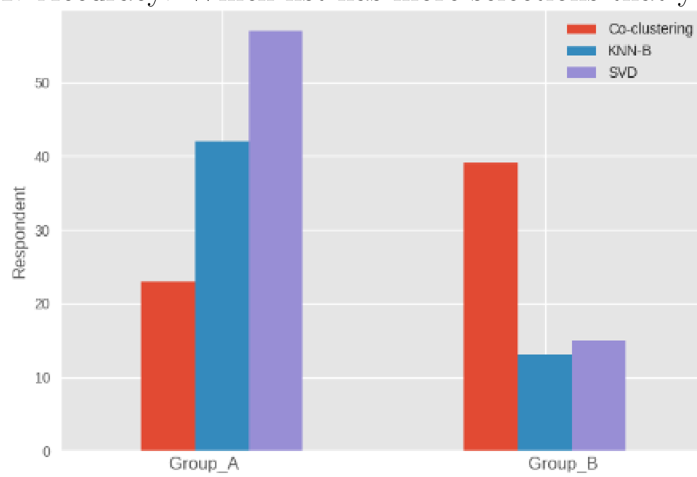
Personality Question	Result
1- I see myself as open to experience, imaginative.	1. Accuracy: Which list has more selections that you find appealing ?  2- Accuracy: Which list has more obviously bad suggestions for you ? 

2. I see myself as dependable, organized.	1. Accuracy: Which list has more selections that you find appealing ?  2- Accuracy: Which list has more obviously bad suggestions for you ? 

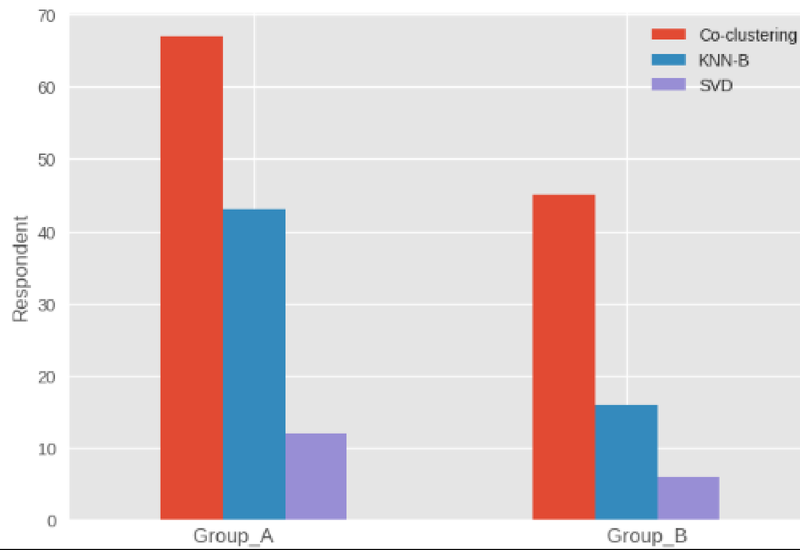
<table>
<tr>
<td>3. I see myself as extroverted, enthusiastic</td>
<td>1. Accuracy: Which list has more selections that you find appealing ?


2- Accuracy: Which list has more obviously bad suggestions for you ?

</td>
</tr>
</table>

<table>
<tr>
<td>4. I see myself as agreeable, kind.</td>
<td>
1. Accuracy: Which list has more selections that you find appealing ?



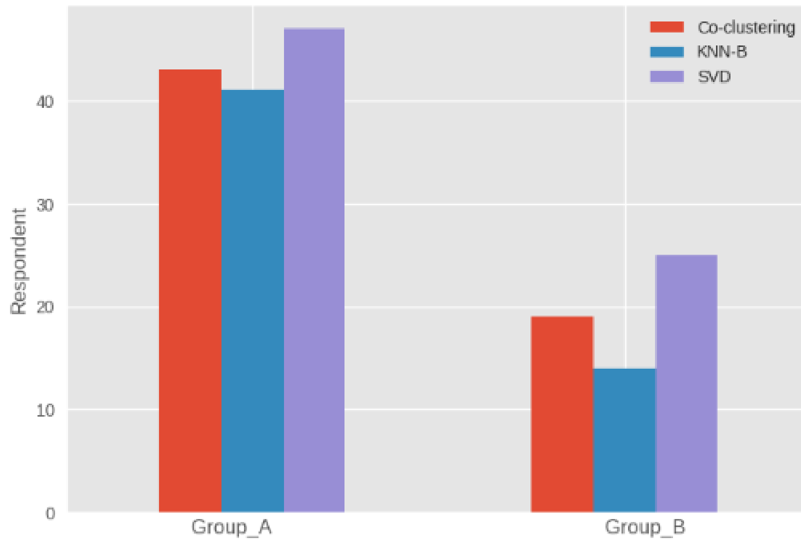
2- Accuracy: Which list has more obviously bad suggestions for you ?



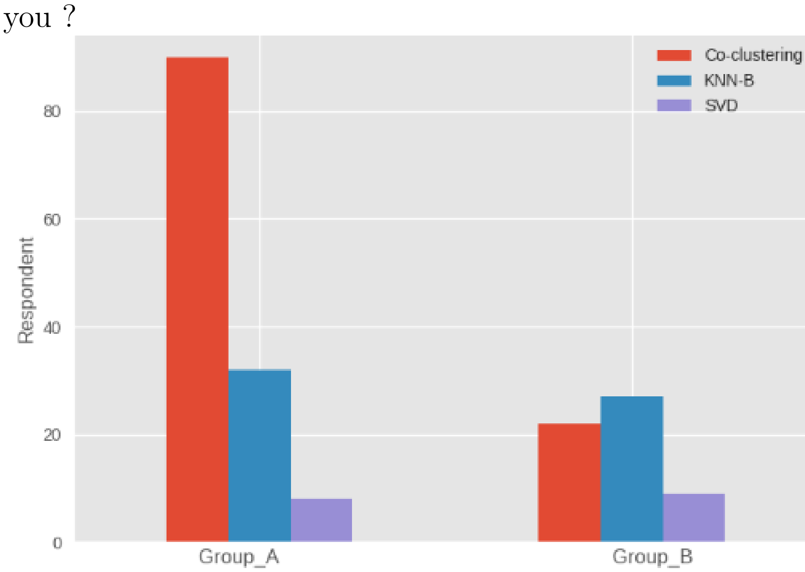
</td>
</tr>
</table>

5. I see myself as emotionally stable, calm

1. Accuracy: Which list has more selections that you find appealing ?



2- Accuracy: Which list has more obviously bad suggestions for you ?



The Table 4.22 shows the overall majority choice per each personality group.

Table 4.22: The overall majority choice per each personality and Accuracy metric.

Personality trait	**Group A**	**Group B**
Openness	SVD	SVD
Conscientiousness	SVD	SVD
Extroversion	SVD	Co-clustering
Agreeableness	SVD	SVD
Neuroticism	SVD	SVD

4.2.3.2 Diversity Metric

Table 4.22 show the differences in participants assessment of the recommendations in terms of diversity, the participant who consider themselves imaginative, open to experience and those who not, thought that SVD generated more similar suggestions, while both groups A and B, considered that KNN-B and Co-clustering algorithms suggested more varied selections than SVD. As for matching the variety of preferences, group A considered Co-clustering to preform worse and and group B considered SVD to perform well.

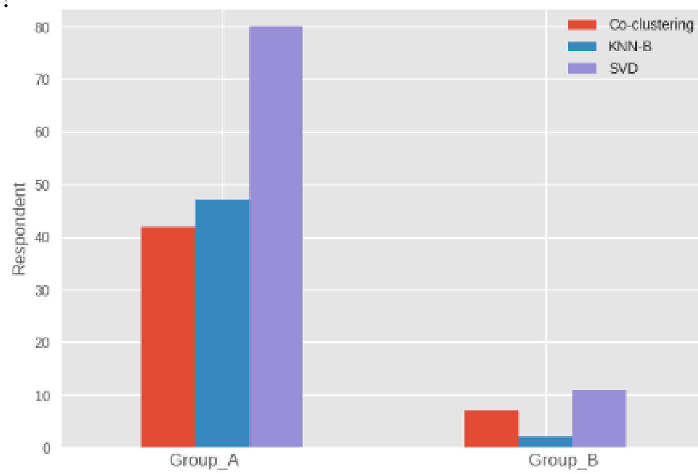
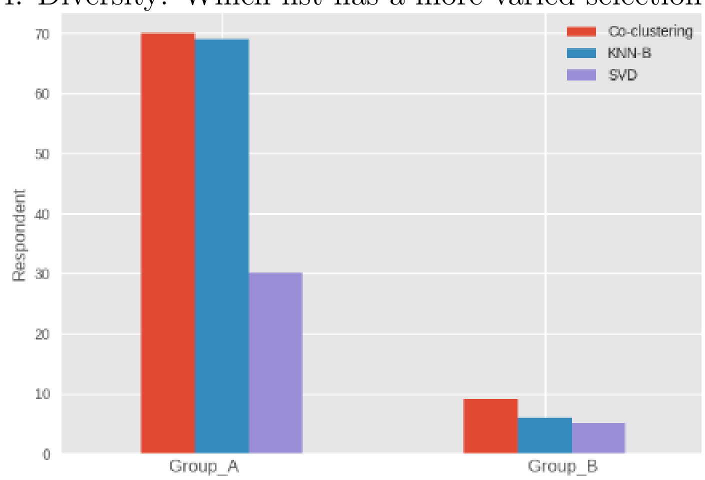
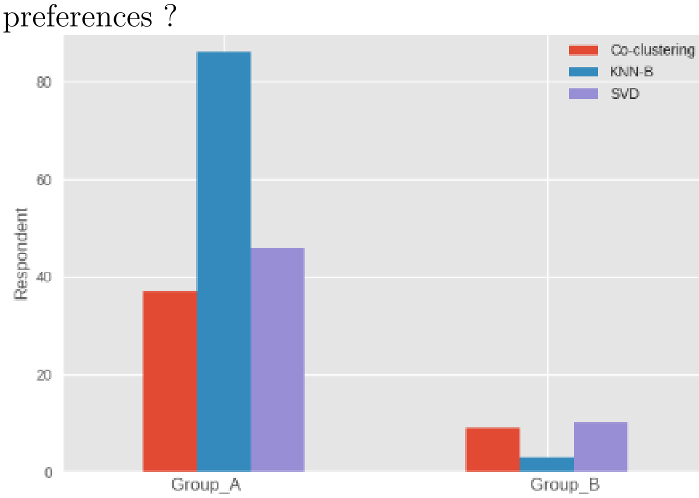
As for the people who consider themselves dependable and organized both groups preferred SVD in terms of providing similar suggestions. For the varied selections Co-clustering and KNN-B were the most preferred by group A participants. In addition, KNN-B highly considered by group A, in terms of suggestions that match wider variety of preferences.

As for participant who consider themselves enthusiastic, they preferred KNN-B and Co-clustering that SVD in terms of match wider variety of preferences, for group B Co-clustering more preferred in terms to generate varied selections.

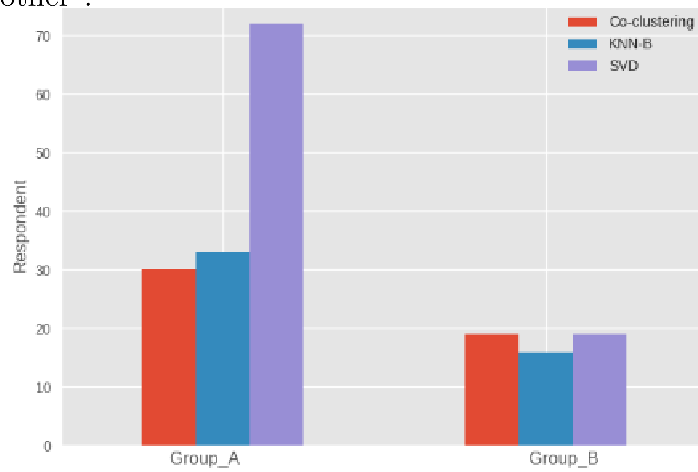
For those who consider themselves kind, both groups they preferred SVD than Co-clustering in terms of generate similar recommendations. For the group B, KNN-B much more preferred in terms of match wider variety of preferences.

Those participants who consider themselves emotionally stable and calm, the Co-clustering algorithm comes first in terms of the diversity, and for group B the KNN-B was the most selected.

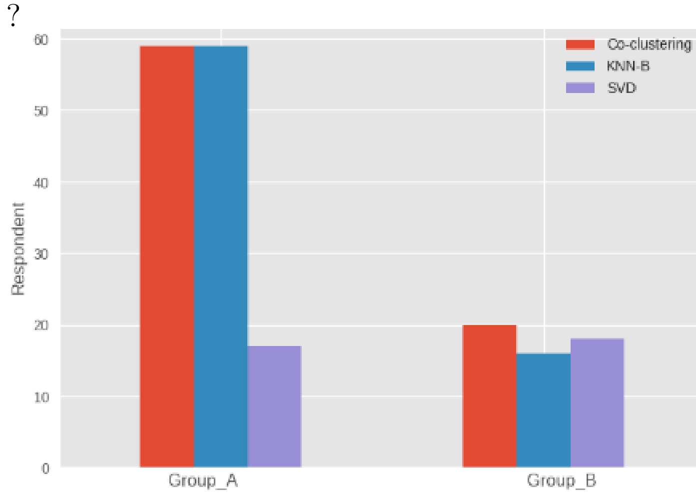
Table 4.23: Diversity and Personality

Personality Question	Result
1- I see myself as open to experience, imaginative.	3. Diversity: Which list has more countries that are similar to each other ?  4. Diversity: Which list has a more varied selection of countries ?  5. Diversity: Which list has countries that match a wider variety of preferences ? 

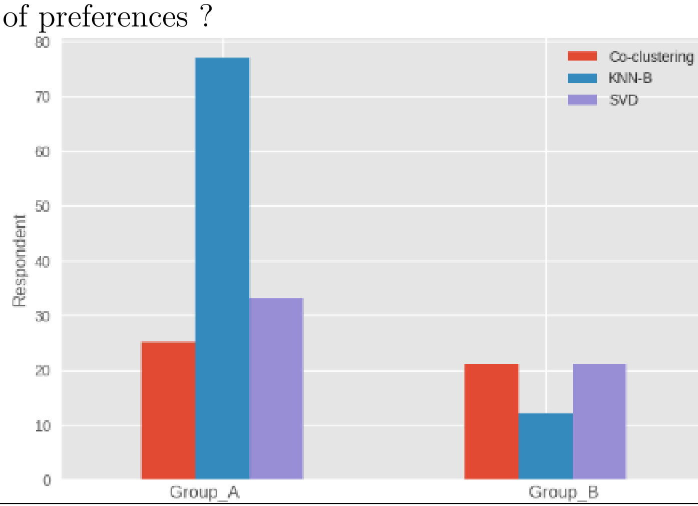
<table>
<tr>
<td>2. I see myself as dependable, organized.</td>
<td>
3. Diversity: Which list has more countries that are similar to each other ?



4. Diversity: Which list has a more varied selection of countries ?

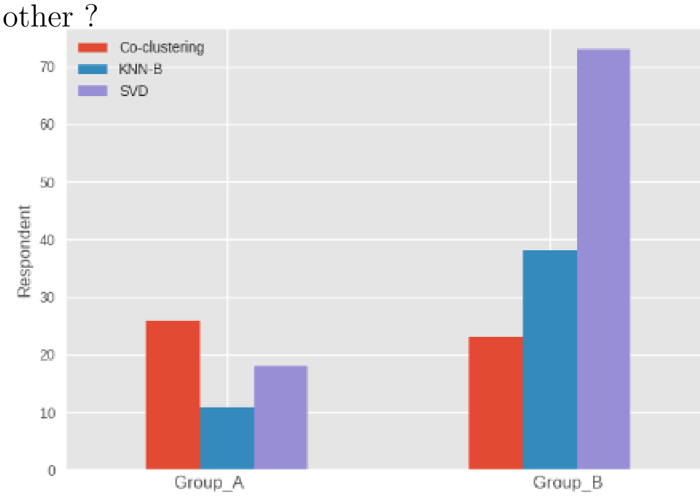


5. Diversity: Which list has countries that match a wider variety of preferences ?

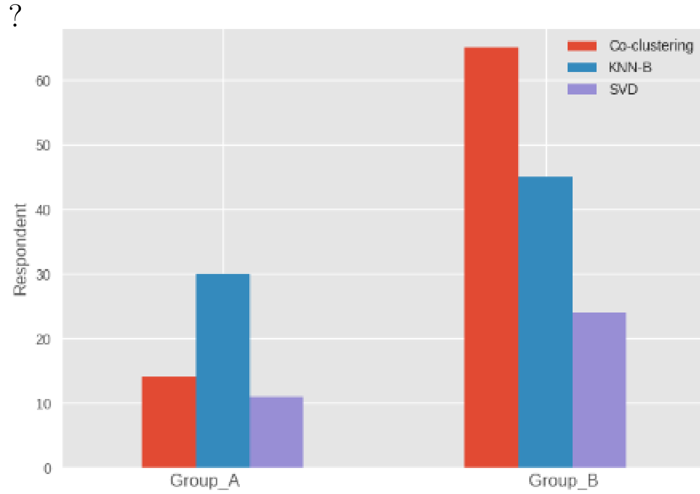


</td>
</tr>
</table>

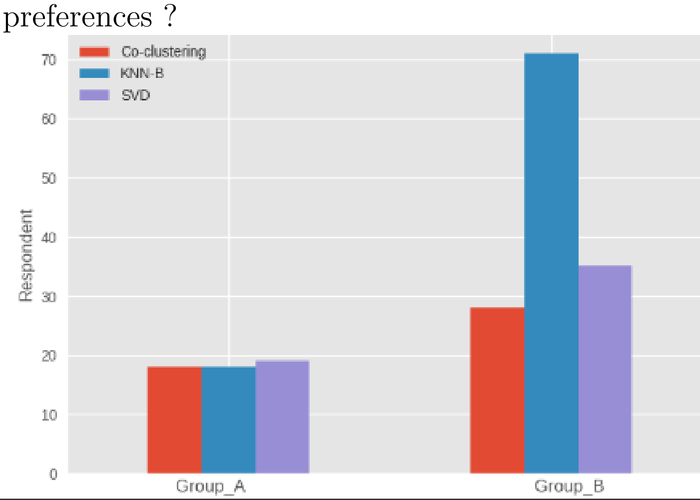
<table>
<tr>
<td>3. I see myself as extroverted, enthusiastic</td>
<td>
3. Diversity: Which list has more countries that are similar to each other ?



4. Diversity: Which list has a more varied selection of countries ?



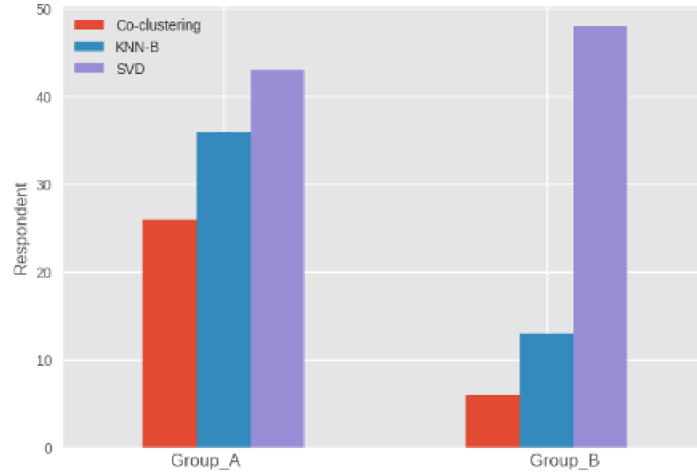
5. Diversity: Which list has countries that match a wider variety of preferences ?



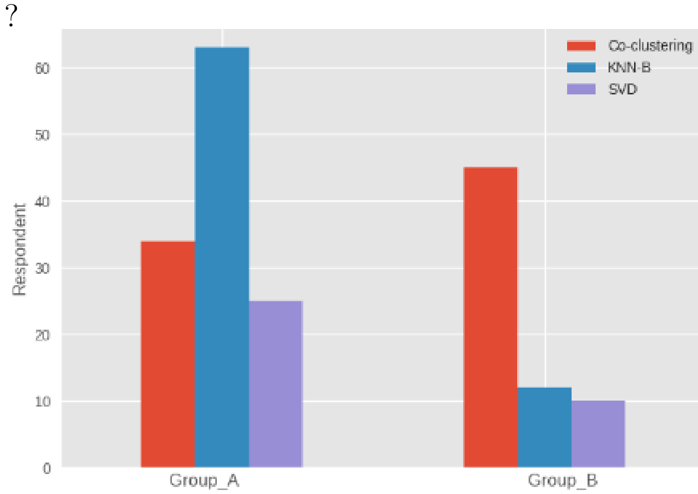
</td>
</tr>
</table>

4. I see myself as agreeable, kind.

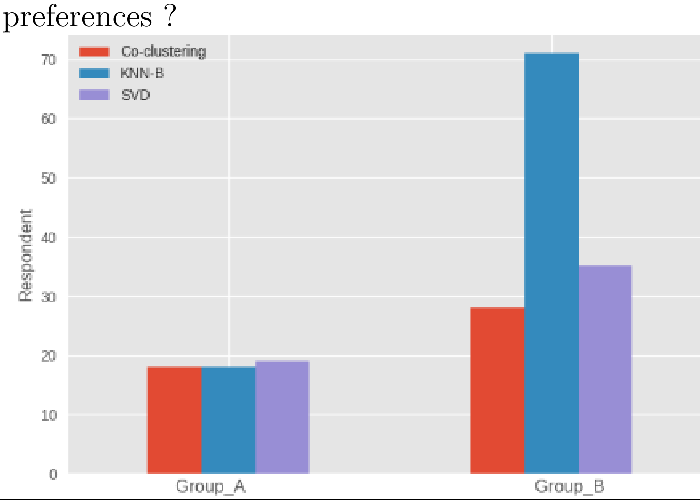
3. Diversity: Which list has more countries that are similar to each other ?



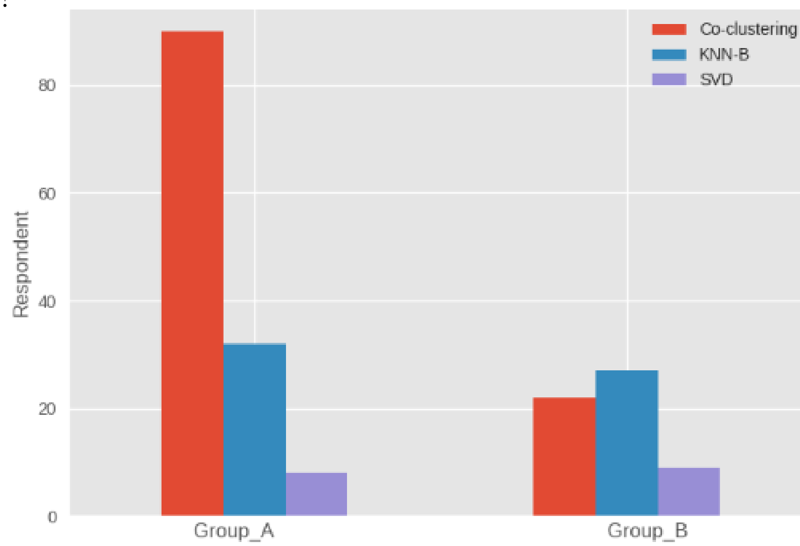
4. Diversity: Which list has a more varied selection of countries ?



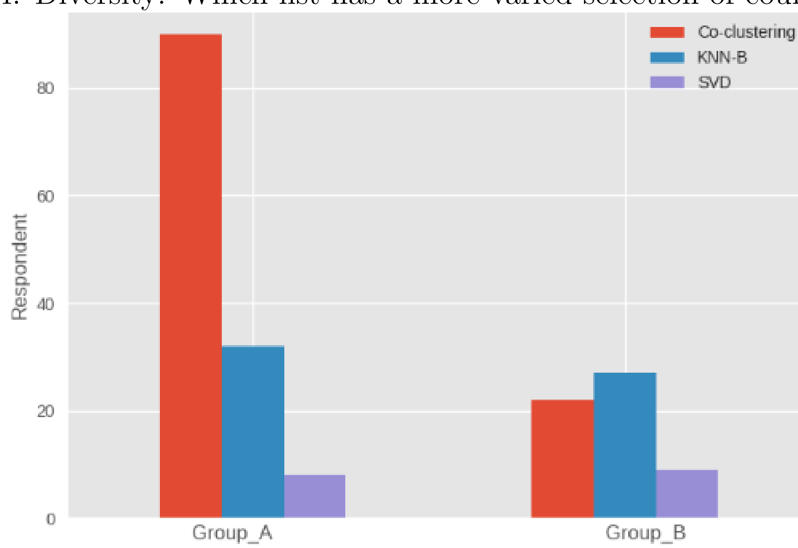
5. Diversity: Which list has countries that match a wider variety of preferences ?



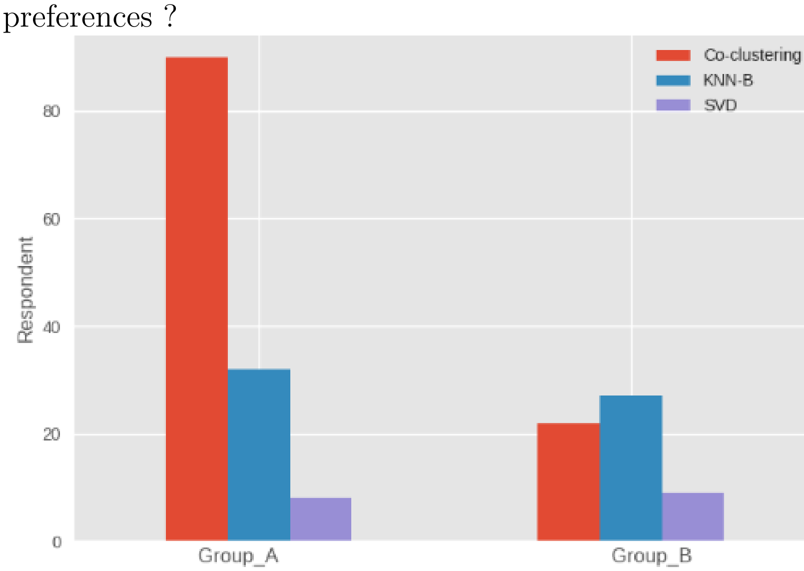
<table>
<tr>
<td>5. I see myself as emotionally stable, calm</td>
<td>
3. Diversity: Which list has more countries that are similar to each other ?



4. Diversity: Which list has a more varied selection of countries ?



5. Diversity: Which list has countries that match a wider variety of preferences ?



</td>
</tr>
</table>

The Table 4.23 shows the overall majority choice per each personality group.

Table 4.24: The overall majority choice per each personality and Diversity metric.

Personality trait	**Group A**	**Group B**
Openness	Co-clustering	Co-clustering
Conscientiousness	Co-clustering	Co-clustering
Extroversion	KNN-B	KNN-B
Agreeableness	KNN-B	Co-clustering
Neuroticism	Co-clustering	KNN-B

4.2.3.3 Understand-Me Metric

Table 4.24 presents the answers to Understand Me metric questions. We can see that the people who consider themselves more open to the experience and imaginative predominantly preferred SVD as better reflecting more personalized recommendations. While those who not considered imaginative, gave few votes to Co-clustering in terms personalizing suggestions and selected SVD as the most generate more mainstreams choices

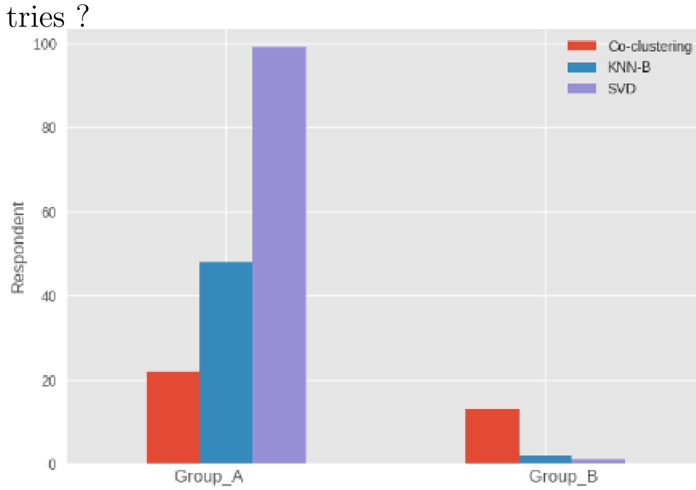
For those who considered themselves organized both groups preferred SVD algorithm as the better reflects their preferences.

Participants who considered themselves extroverted and enthusiastic and those who not, preferred SVD in terms of Understand Me metric.

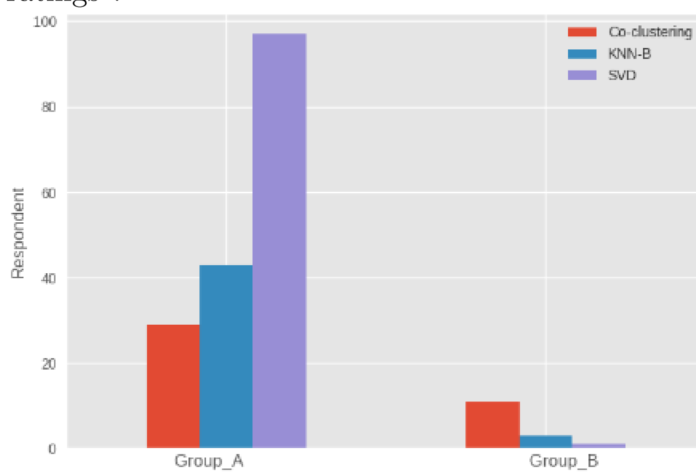
In general the most considered algorithm in tems of Understand Me metric was SVD.

Table 4.25: Understand-Me and Personality

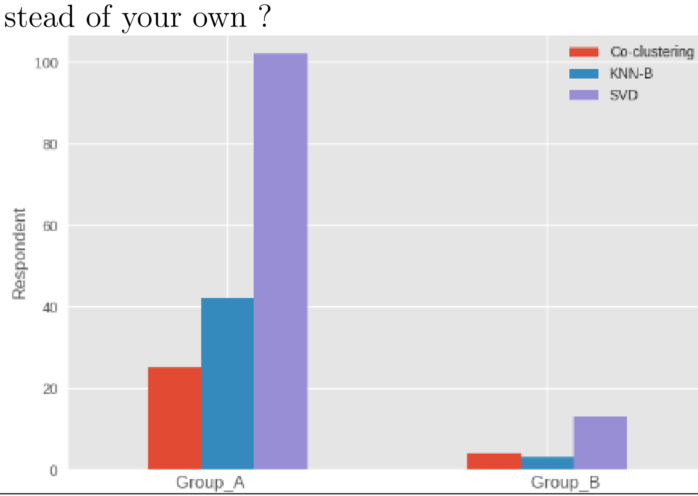
<table>
<tr><th>Personality Question</th><th>Result</th></tr>
<tr><td>1- I see myself as open to experience, imaginative.</td><td>
6. Understand Me : Which list better reflects your preferences in countries ?



7. Understand Me: Which list seems more personalized to your countries ratings ?

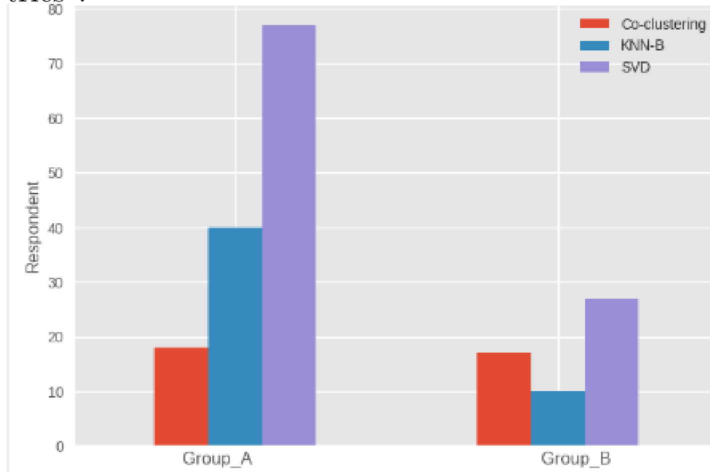


8. Understand Me: Which list represents more mainstream ratings instead of your own ?

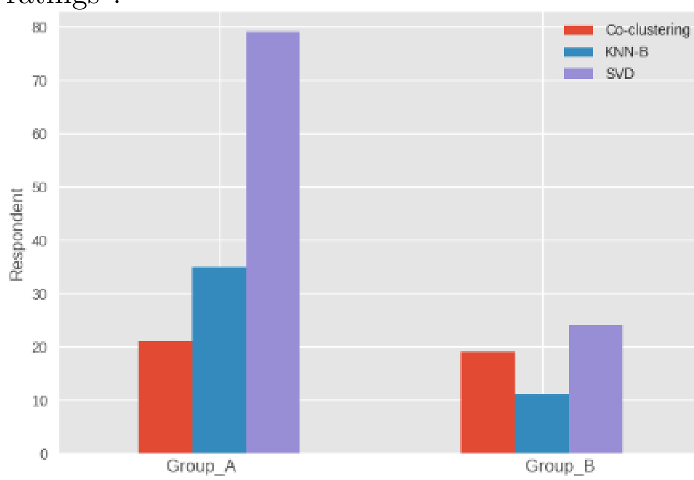


</td></tr>
</table>

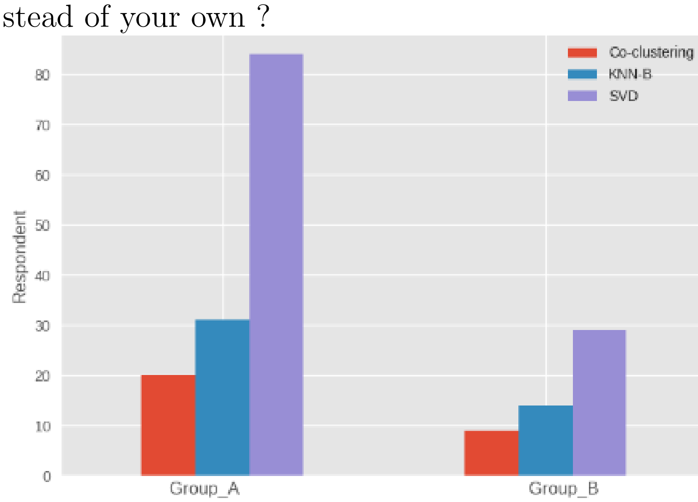
<table>
<tr>
<td>2. I see myself as dependable, organized.</td>
<td>
6. Understand Me : Which list better reflects your preferences in countries ?



7. Understand Me: Which list seems more personalized to your countries ratings ?

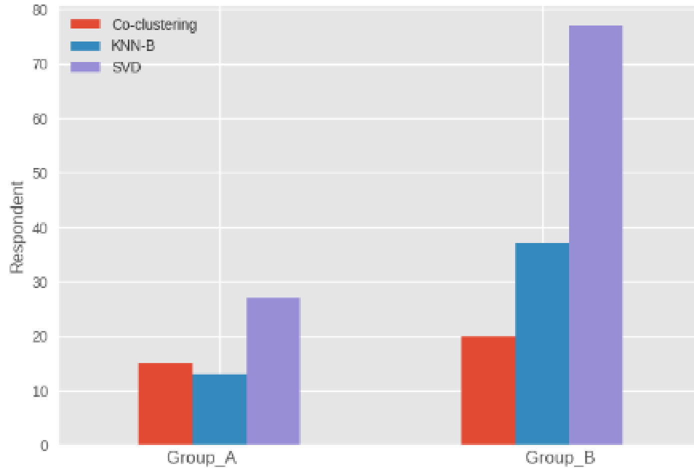


8. Understand Me: Which list represents more mainstream ratings instead of your own ?

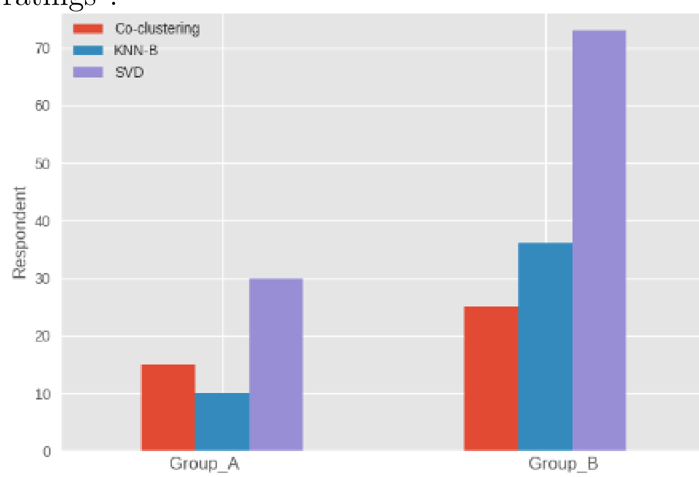


</td>
</tr>
</table>

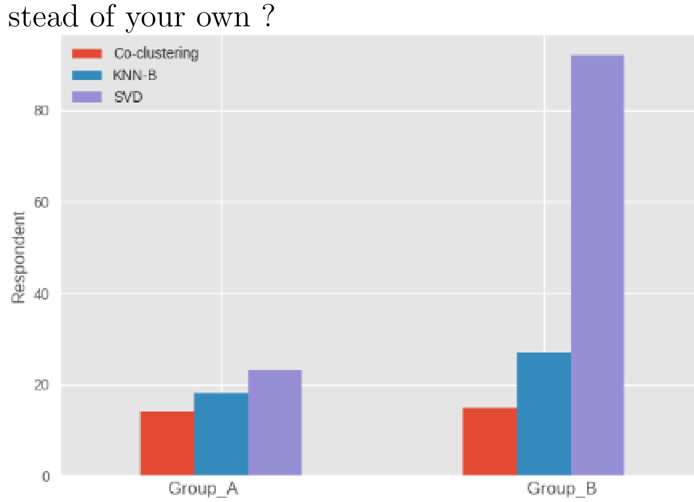
<table>
<tr>
<td>3. I see myself as extroverted, enthusiastic</td>
<td>
6. Understand Me : Which list better reflects your preferences in countries ?



7. Understand Me: Which list seems more personalized to your countries ratings ?

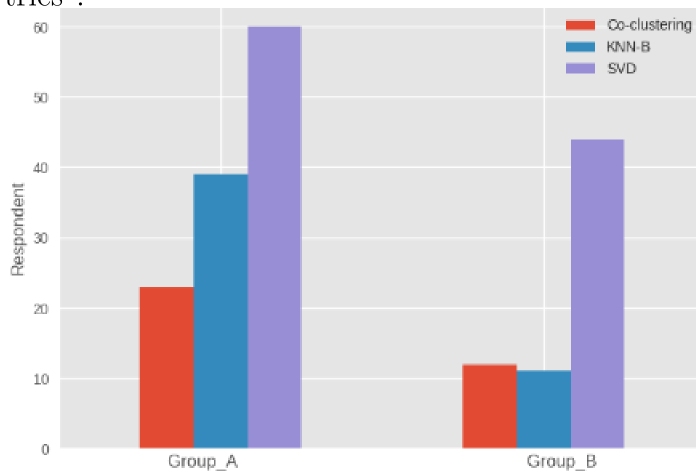


8. Understand Me: Which list represents more mainstream ratings instead of your own ?

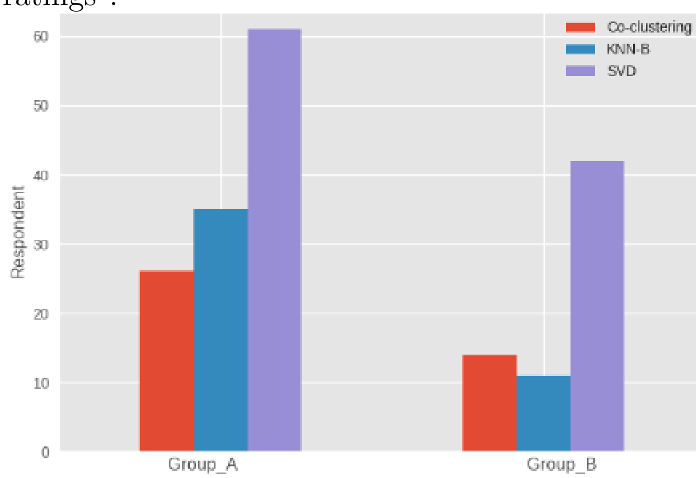


</td>
</tr>
</table>

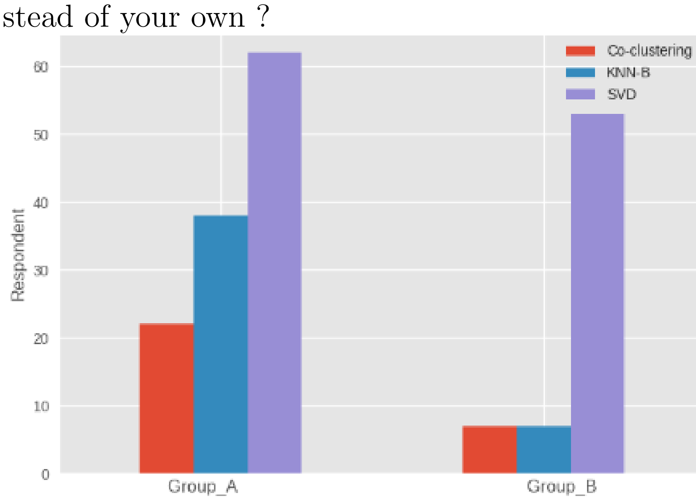
<table>
<tr>
<td>4. I see myself as agreeable, kind.</td>
<td>
6. Understand Me : Which list better reflects your preferences in countries ?



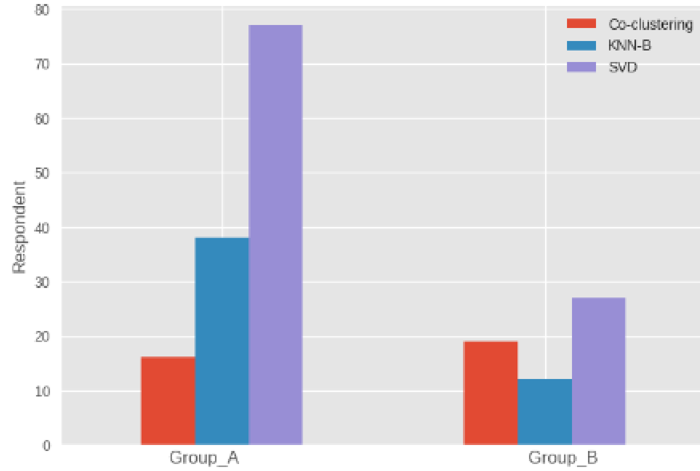
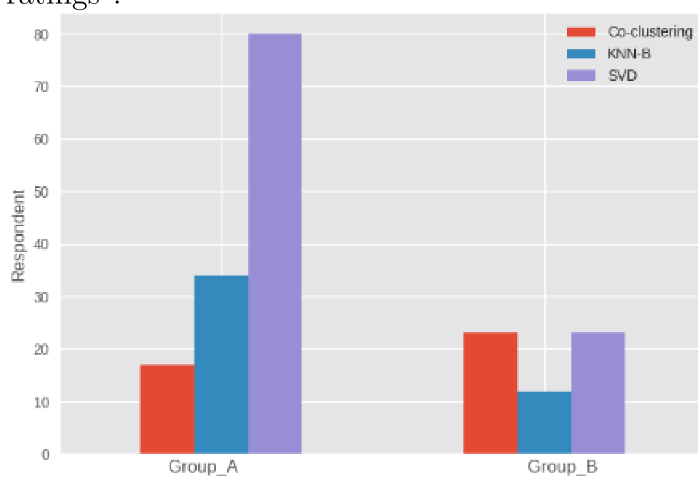
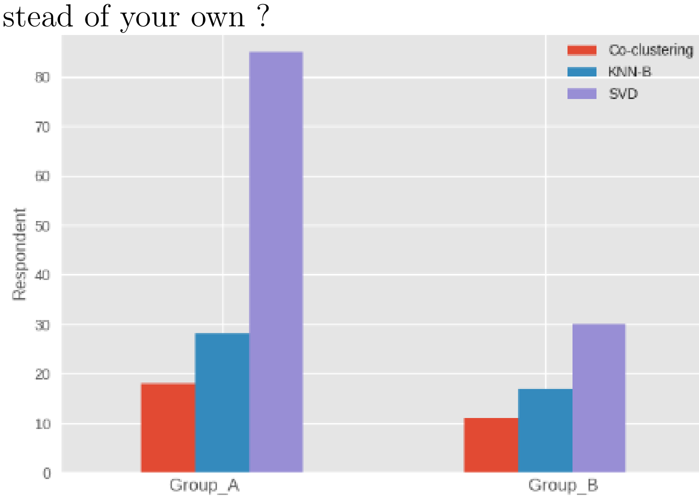
7. Understand Me: Which list seems more personalized to your countries ratings ?



8. Understand Me: Which list represents more mainstream ratings instead of your own ?



</td>
</tr>
</table>

5. I see myself as emotionally stable, calm	6. Understand Me : Which list better reflects your preferences in countries ?  7. Understand Me: Which list seems more personalized to your countries ratings ?  8. Understand Me: Which list represents more mainstream ratings instead of your own ? 

The Table 4.25 shows the overall majority choice per each personality group.

Table 4.26: The overall majority choice per each personality and Understand Me metric

Personality trait	**Group A**	**Group B**
Openness	SVD	Co-clustering
Conscientiousness	SVD	SVD
Extroversion	SVD	SVD
Agreeableness	SVD	SVD
Neuroticism	SVD	SVD

4.2.3.4 Satisfaction Metric

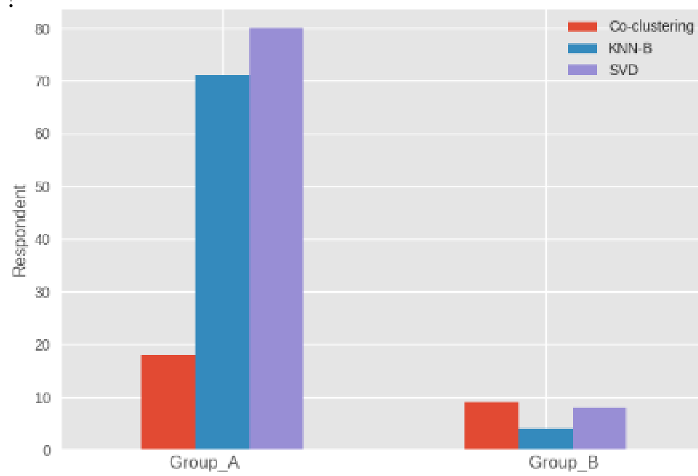
Table 4.27 present the answers to the satisfaction metric questions. Participants who consider themselves open to experience and imaginative, considered SVD as the algorithm that generated the most recommendations. Those who do not consider themselves imaginative considered KNN-B perform worst that Co-clustering and SVD.

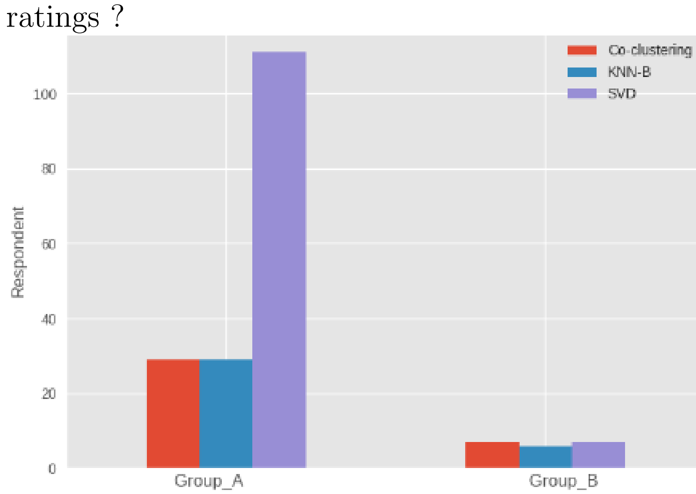
For those respondents who do not consider themselves highly dependable and organized, preferred SVD in terms if satisfaction metric.

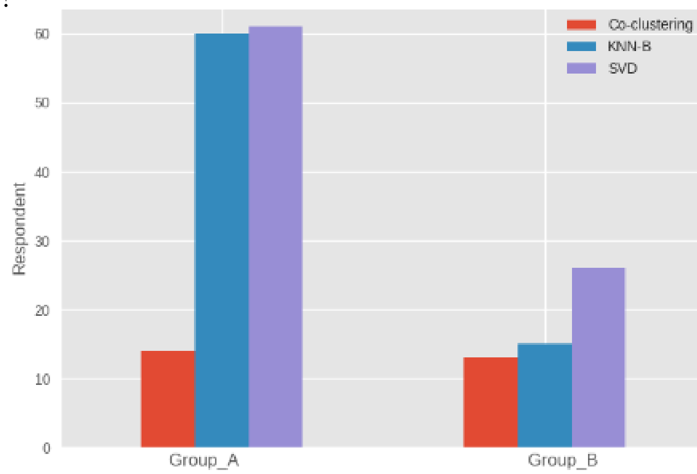
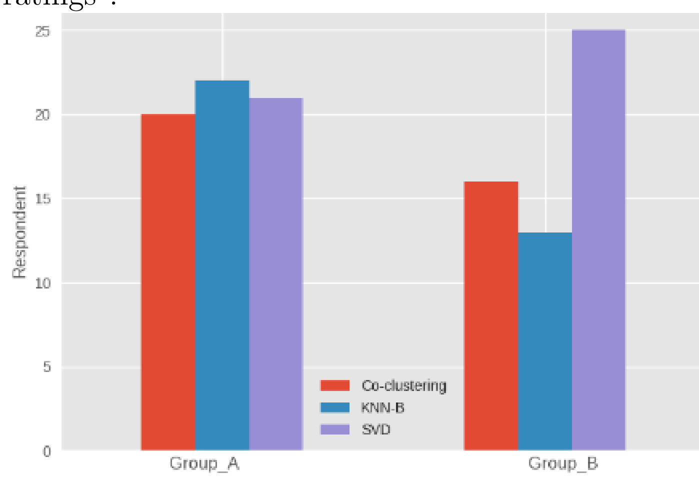
Participants who saw themselves as extroverted and enthusiastic preferred SVD as much more often. For the group B though that Co-clustering was performing worst.

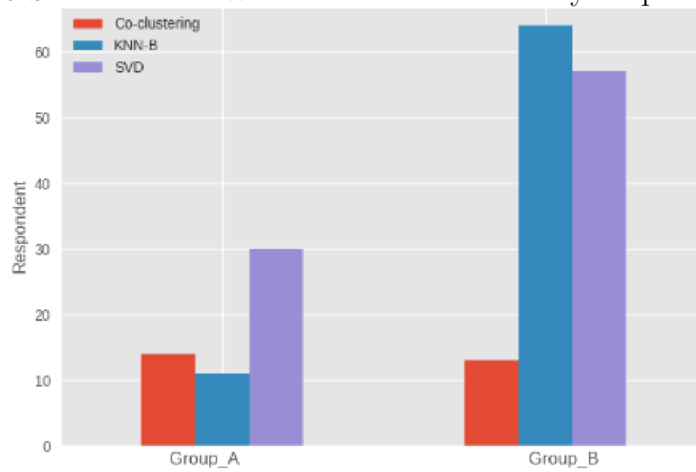
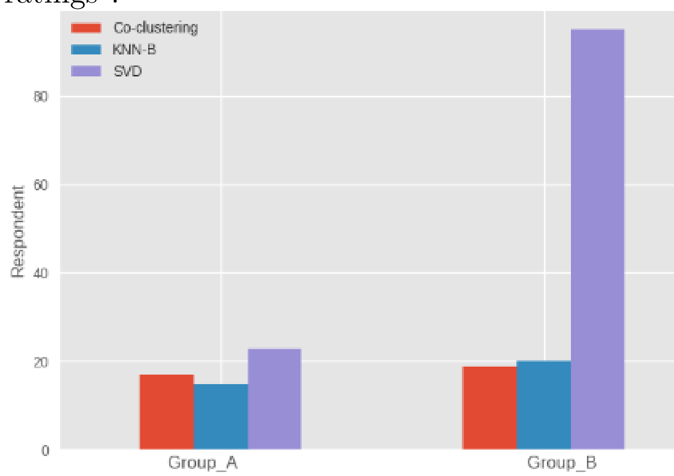
For those consider themselves more kind and emotionally stable and those who not as such, they both preferred SVD in terms of satisfaction metric.

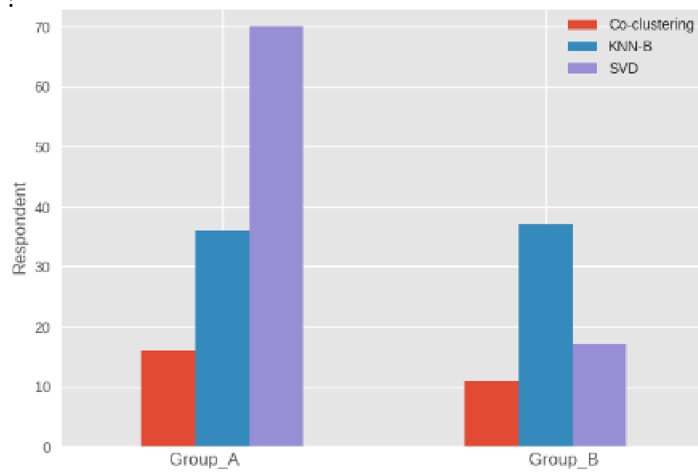
Table 4.27: Satisfaction and Personality

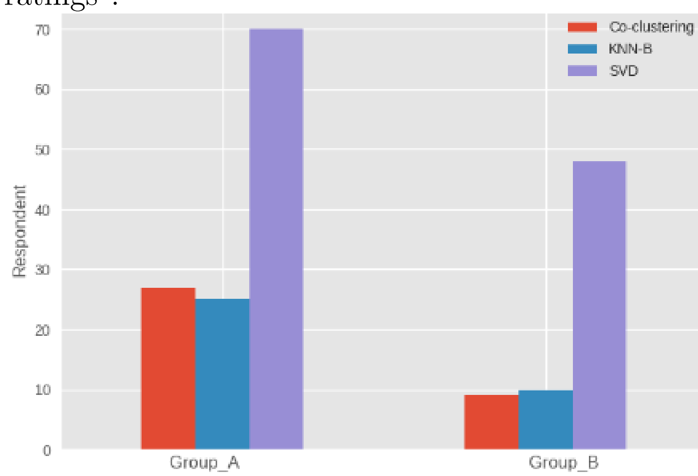
<table>
<tr><th>Personality Question</th><th>Result</th></tr>
<tr><td rowspan="1">1- I see myself as open to experience, imaginative.</td><td>9. Satisfaction : Which list better reflects your preferences in countries ?


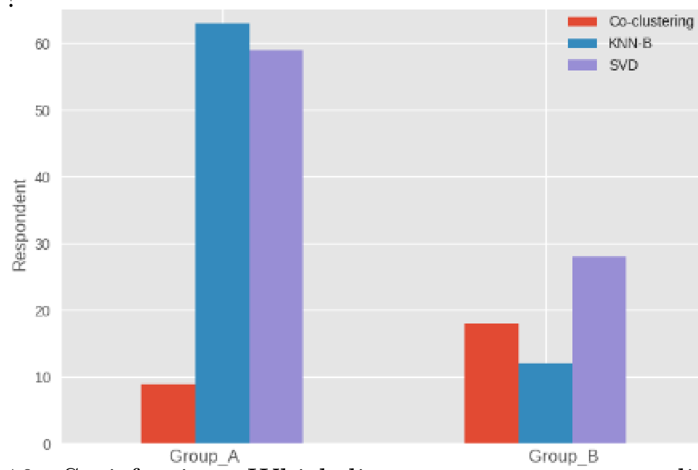
10. Satisfaction: Which list seems more personalized to your countries ratings ?

</td></tr>
</table>

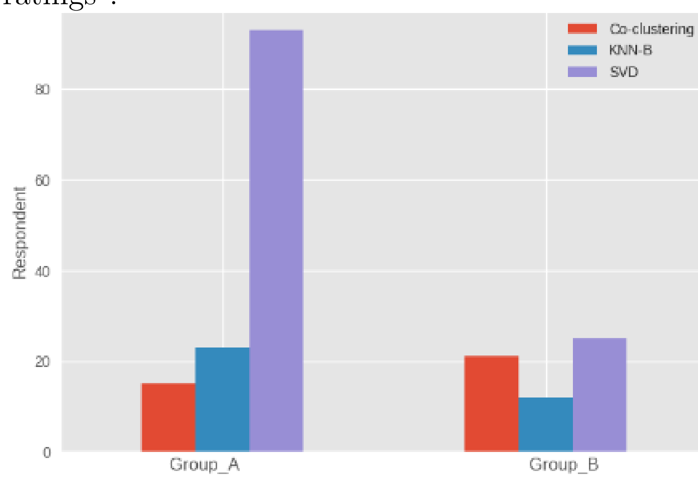
2. I see myself as dependable, organized.	9. Satisfaction : Which list better reflects your preferences in countries ?  10. Satisfaction: Which list seems more personalized to your countries ratings ? 

3. I see myself as extroverted, enthusiastic	9.Satisfaction : Which list better reflects your preferences in countries ?  10.Satisfaction: Which list seems more personalized to your countries ratings ? 

<table>
<tr>
<td>4. I see myself as agreeable, kind.</td>
<td>9. Satisfaction : Which list better reflects your preferences in countries ?


10. Satisfaction: Which list seems more personalized to your countries ratings ?

</td>
</tr>
</table>

<table>
<tr>
<td>5. I see myself as emotionally stable, calm</td>
<td>9. Satisfaction : Which list better reflects your preferences in countries ?


10. Satisfaction: Which list seems more personalized to your countries ratings ?


</td>
</tr>
</table>

the Table 4.28 shows the overall majority choice per each personality group.

Table 4.28: The overall majority choice per each personality and Satisfaction metric.

Personality trait	**Group A**	**Group B**
Openness	SVD	Co-Clustering
Conscientiousness	SVD	SVD
Extroversion	SVD	SVD
Agreeableness	SVD	SVD
Neuroticism	SVD	SVD

4.2.3.5 Novelty Metric

Table 4.29 presents the answers to the Novelty metric, for those who consider themselves open to experience and imaginative, thought than KNN-B produced pleasantly surprising suggestions that Co-clustering, SVD was the most considered in terms of producing fewer suggestions.

Those who see themselves organized and dependable and those who not they both preferred SVD in terms of produce fewer new suggestions.

Both groups A and B, who consider themselves enthusiastic and those who not thought that KNN-B produce more pleasantly surprising countries.

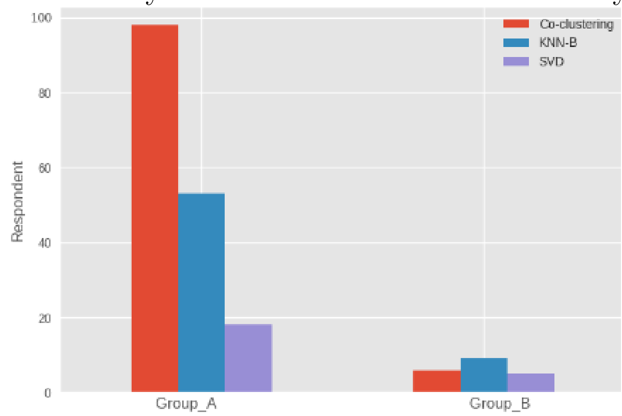
For those who consider themselves kind and those who not they preferred KNN-B as the most one generates more pleasantly surprising and SVD considered the one that generates fewer new suggestions.

Participants who consider themselves emotionally stable claimed that list produced by KNN-B more pleasantly surprising, and those who not preferred list produced by Co-clustering, while SVD considered by both groups in terms of provides fewer new suggestions.

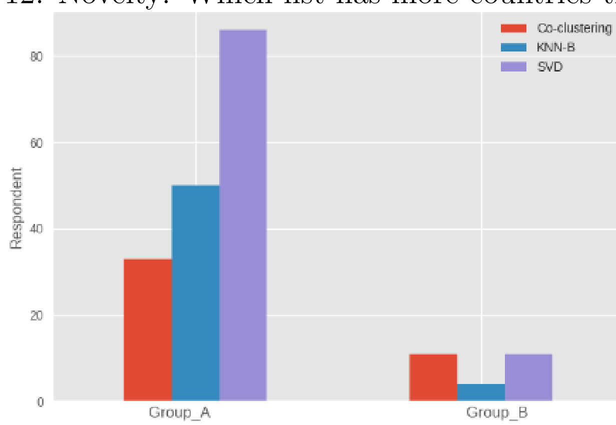
Table 4.29: Novelty and Personality

<table>
<tr><th>Personality Question</th><th>Result</th></tr>
<tr>
<td>1- I see myself as open to experience, imaginative.</td>
<td>

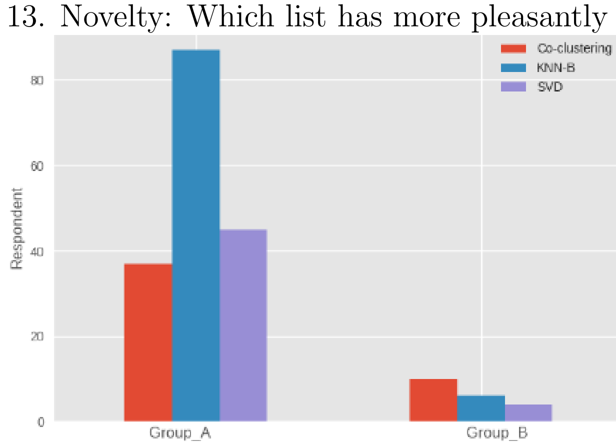
11. Novelty : Which list has more countries you did not expect ?



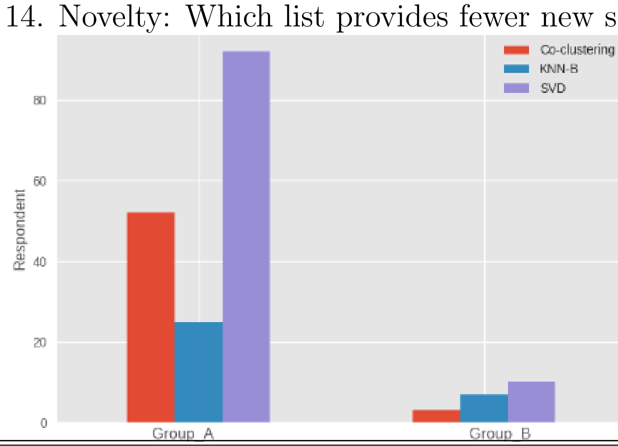
12. Novelty: Which list has more countries that are familiar to you ?



13. Novelty: Which list has more pleasantly surprising countries ?



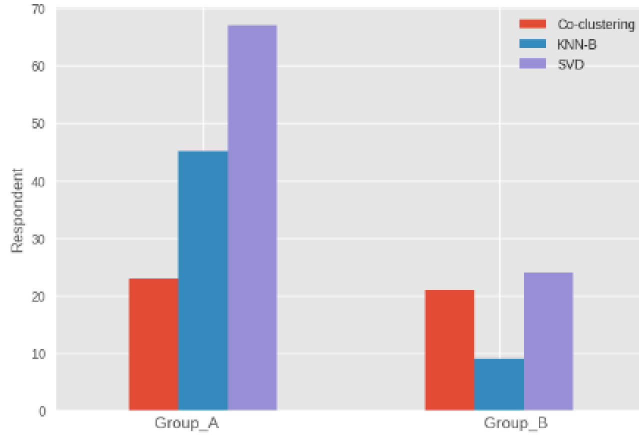
14. Novelty: Which list provides fewer new suggestions ?



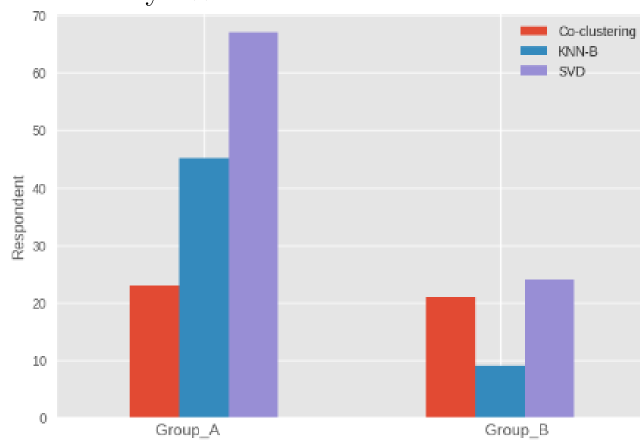
</td>
</tr>
</table>

2. I see myself as dependable, organized.

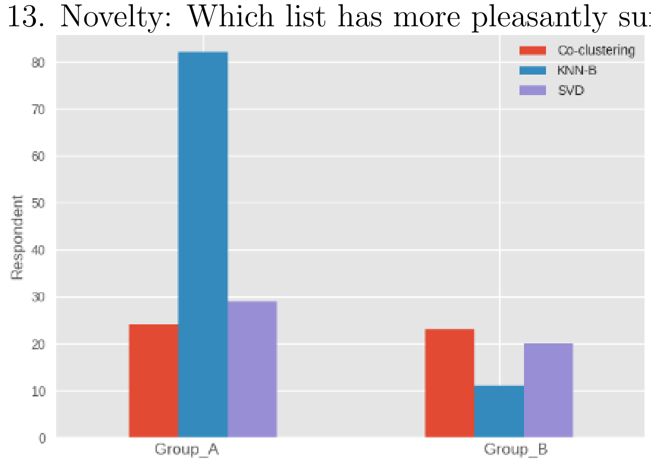
11. Novelty : Which list has more countries you did not expect ?



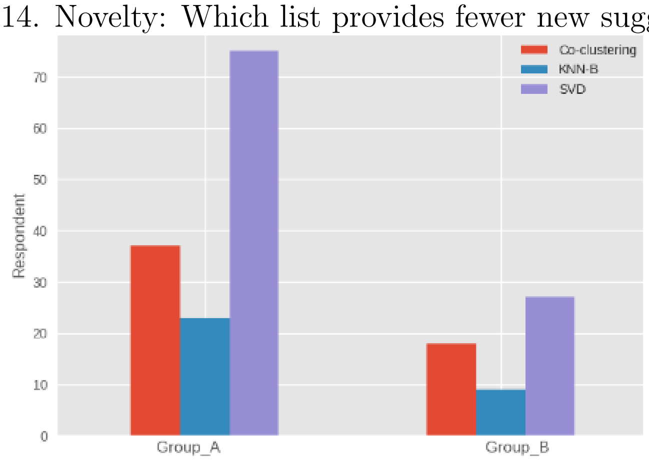
12. Novelty: Which list has more countries that are familiar to you ?



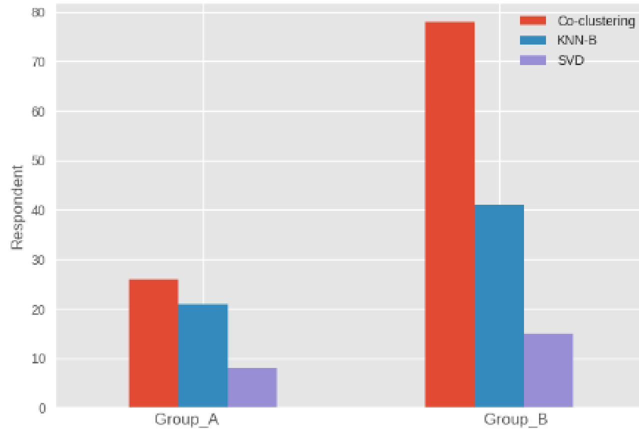
13. Novelty: Which list has more pleasantly surprising countries ?



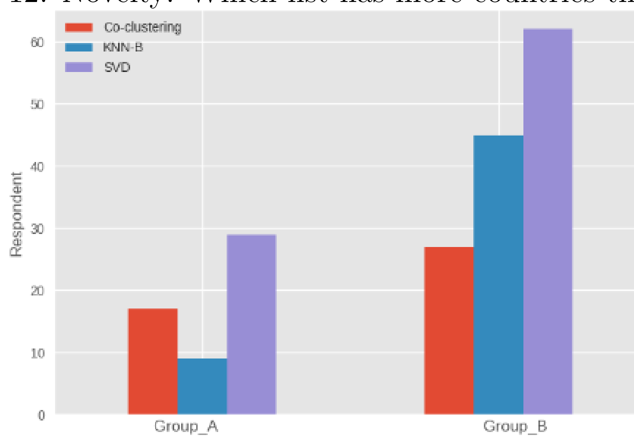
14. Novelty: Which list provides fewer new suggestions ?



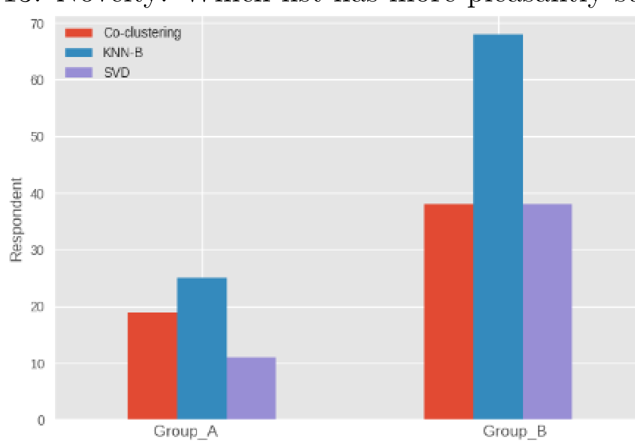
<table>
<tr>
<td>3. I see myself as extroverted, enthusiastic</td>
<td>
11. Novelty : Which list has more countries you did not expect ?



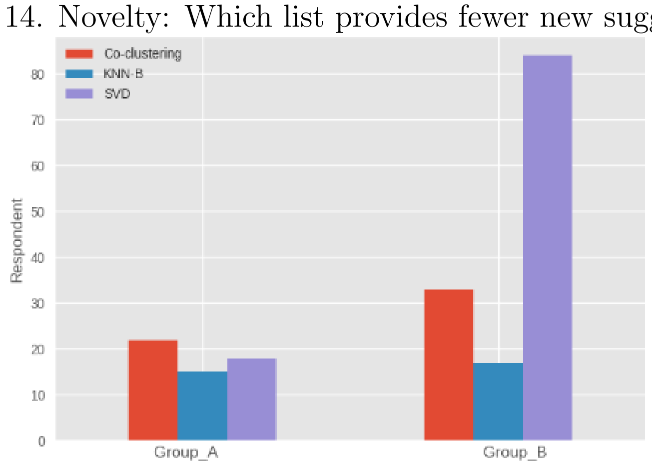
12. Novelty: Which list has more countries that are familiar to you ?



13. Novelty: Which list has more pleasantly surprising countries ?

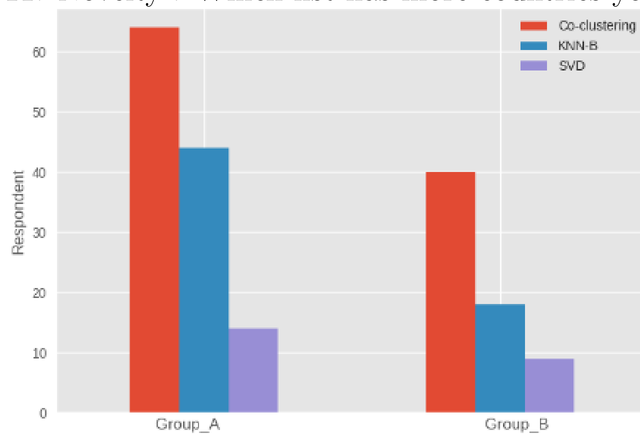


14. Novelty: Which list provides fewer new suggestions ?

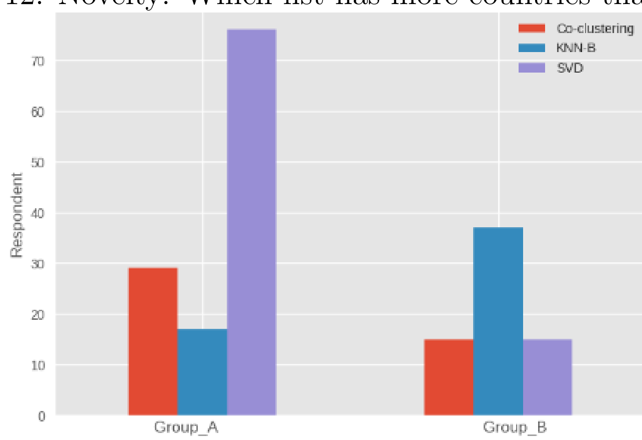


</td>
</tr>
</table>

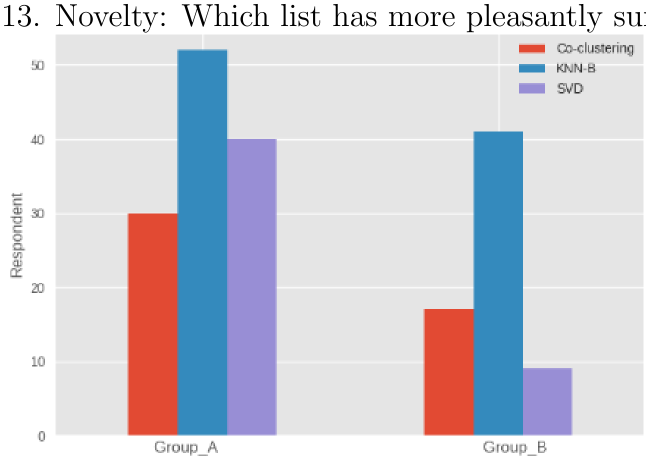
<table>
<tr>
<td>4. I see myself as agreeable, kind.</td>
<td>
11. Novelty : Which list has more countries you did not expect ?



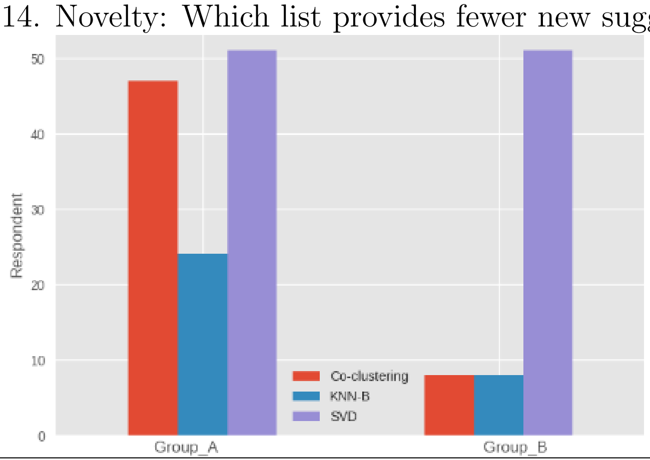
12. Novelty: Which list has more countries that are familiar to you ?



13. Novelty: Which list has more pleasantly surprising countries ?



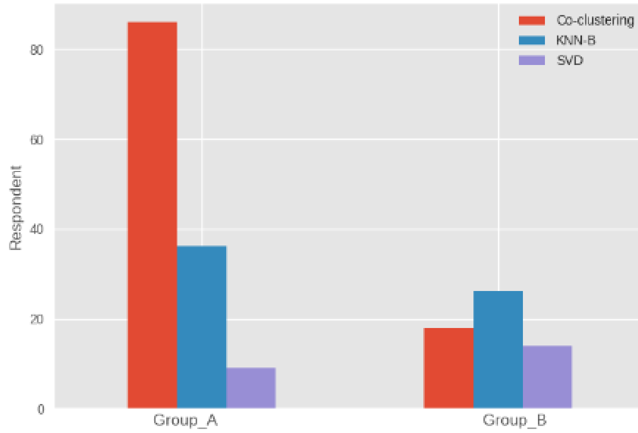
14. Novelty: Which list provides fewer new suggestions ?



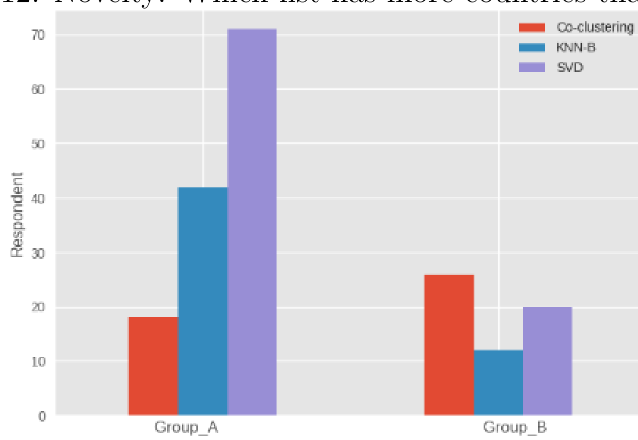
</td>
</tr>
</table>

5. I see myself as emotionally stable, calm

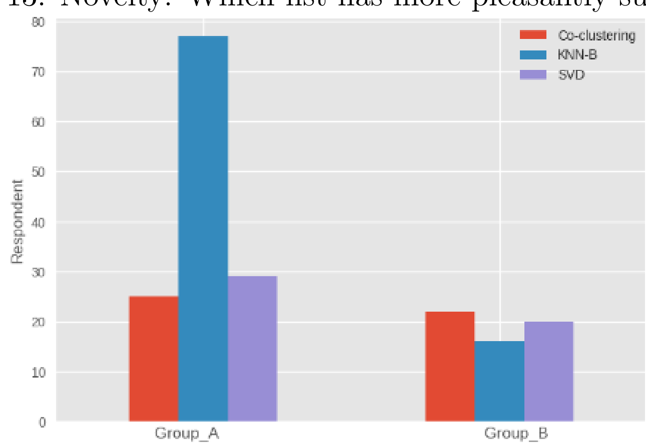
11. Novelty : Which list has more countries you did not expect ? ?



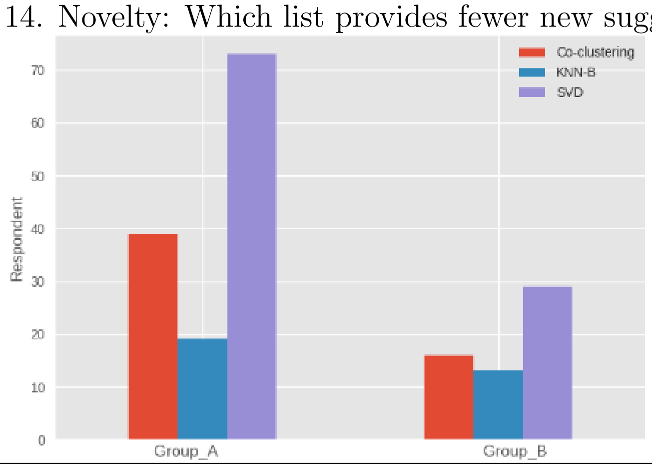
12. Novelty: Which list has more countries that are familiar to you ?



13. Novelty: Which list has more pleasantly surprising countries ?



14. Novelty: Which list provides fewer new suggestions ?



The Table 4.28 shows the overall majority choice per each personality group.

Table 4.30: The overall majority choice per each personality and Novelty metric

Personality trait	**Group A**	**Group B**
Openness	KNN-B	Co-clustering
Conscientiousness	KNN-B	KNN-B
Extroversion	KNN-B	KNN-B
Agreeableness	KNN-B	KNN-B
Neuroticism	KNN-B	Co-clustering

4.2.4 The system Usability

At the end of the interaction with the system we requested our participants to assess their experience in terms of usability. We used the well-know system Usability Survey (SUS) [79]. Is a ten-item questionnaire based on five-point Likert scale 2.2. A SUS score above 68 is considered to be above average [82]. The general guideline of the SUS score interpretation is presented in Table 4.31, this data was adopted from [83].

Table 4.31: SUS Score Interpretation

SUS Score	**Grade**	**Adjective Rating**
>80	A	Excellent
68 – 80.3	B	Good
68	C	Okay
51 – 68	D	Poor
< 51	F	Awful

The result obtained from usability anaylsis show that our participants have assessed our system less than our expectations, in other words less than "Okey" as adjective rating. Indeed, while actual values given by users ranged from 22.5 to 100, the mean was 60 and median score of 62. The table below represents the overall scores.

Table 4.32: SUS Survey Responses

Question	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. I think that I would like to use this system frequently	21%	26%	36%	16%	1%
2. I found the system unnecessary complex	12%	40%	26%	16%	6%
3. I thought the system was easy to use	5%	11%	20%	16%	48%
4. I think that I would need the support of a technical person to be able to use this system	50%	36%	9%	5%	0%
5. I found the various functions of this system were well integrated	6%	13%	43%	33%	5%
6. I thought there was too much inconsistency in this system	5%	30%	34%	22%	8%
7. I would imagine that most people would learn to use this system very quickly	1%	8%	24%	48%	19%
8. I found the system very cumbersome (inconvenient)	11%	39%	27%	17%	6%
9. I felt very confident using the system	5%	13%	40%	32%	10%
10. I needed to learn a lot of things before I could get going with this system	36%	41%	15%	7%	1%

The final score we calculated (60.82), indicated that we assign our system a grade of D and an adjective rating of "Poor". The SUS calculation revealed to us that the system is on the lower side of "Okey" rather than the higher one. To investigate further, we looked at the comments of the users. Most of the users complained about the number of questions they needed to answer during the survey, which makes it hard for most to complete all the tasks. Others claimed that we should add more questions about peoples' preferences in geography, climate, language type and so on. In addition, most were seeking ratings about the recommended list, and claimed that the lack of knowledge about most of the countries made it difficult to even undertake a task of ratings.

4.3 Conclusion

In this chapter, we gave a detailed description of the training dataset and the source of the features that users used to decide where to emigrate. After that we analyzed the results and made some observations. The Next chapter will be dedicated for the conclusions and future works. The source code illustrated in this part can be found in [84].

Chapter 5

Conclusion

5.1 Summary

Recommender system witnessed a significant progress over the last decade in many fields. In this thesis, we investigate the utility of recommender system algorithms in a new domain. The idea was, developing and deploying a country recommender system, that can overcome the challenge that faces people, when it comes to making the decision to move to another country. In the first chapter, we introduced the problem and presented the research question that we addressee:

- **RQ1 :** Which recommender algorithms can be adopted -based on the preferences of users in order to generate personalized country ranking ?
- **RQ2 :** What are the most important features that users consider when deciding to move to another country ?
- **RQ3 :** Do recommender algorithm preferences depend on personality types ?
- **RQ4 :** Will the system for generating personalized country ranking be usable according to the user's assessment ?

In the second chapter, we presented the needed vocabularies, different recommender systems approaches and we discussed the main relation between user decision making and the personality, while the third dedicated to the proposed methodology. Chapter four started by the details about the set of experiments that we conducted, then the results and analysis were presented in the second section of the chapter, those results can be summarized ass follows.

- **RQ1 :**

  Through our offline validation test, we found that the considered algorithms can give better results in terms of users' assessments. The SVD algorithm performs much better than Co-clustering and KNN-B in terms of accuracy, while KNN-B and Co-clustering gave better results than SVD in terms of diversity. In terms of Understand Me metric SVD was found to suggest the most personalized items, while in terms of novelty, users found SVD to be the worst performing.

- **RQ2 :**

  WE have observed that the majority of participants concerned with *Work opportunities*, *Education quality* and *Health care* when deciding to emigrate to another country. While most of them were not much concerned about social relations such as *Family member abroad* and *Shorter distance*. In addition people who considered themselves imaginative gave much more attention to work opportunities than those do not, see health care much more important than work opportunities, although most of other users with other personalities, concerned with *Work opportunities, Education quality* and health care come first.

- **RQ3 :**

  The results roved that people with different types of personalities tend to choose different type of algorithms. For example, the participants who considered themselves enthusiastic claimed that SVD was more accurate than Co-clustering, and those who do not , claimed that Co-clustering was accurate than SVD and and even KNN-B.

- **RQ4 :**

  Finally, the proposed system has been found to score a bit low in the SUS score. We classified it's usability as "poor" and currently thinking about redesigning it in order to enhance the user experience.

5.2 Future works

As future work direction, we would like to investigate whether RS based on deep learning techniques would improve the quality of recommendations. In addition, we also see the benefit to make our recommendations related to the user's preferences (immigration factors) in order to make the system more personalized, and better tackle the cold start problem.
Finally as a result of the relation between the human personality and algorithms selection, their is a need to incorporate the personality information in the prediction model.

Bibliography

[1] A. Koene, E. Perez, C. J. Carter, R. Statache, S. Adolphs, C. O'Malley, T. Rodden, and D. McAuley, "Ethics of personalized information filtering," in *International Conference on Internet Science*. Springer, 2015, pp. 123–132.

[2] F. Castelli, "Drivers of migration: why do people move?" *Journal of travel medicine*, vol. 25, no. 1, p. tay040, 2018.

[3] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*. Springer, 2011, pp. 1–35.

[4] C. Anderson, *The Long Tail. Random House Business*, 2006.

[5] D. Gavalas, V. Kasapakis, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "A survey on mobile tourism recommender systems," in *2013 Third International Conference on Communications and Information Technology (ICCIT)*. IEEE, 2013, pp. 131–135.

[6] D. Cavada, M. Elahi, D. Massimo, S. Maule, E. Not, F. Ricci, and A. Venturini, "Tangible tourism with the internet of things," in *Information and Communication Technologies in Tourism 2018*. Springer, 2018, pp. 349–361.

[7] J. F. McCarthy, "Pocket restaurantfinder: A situated recommender system for groups," in *Workshop on Mobile Ad-Hoc Communication at the 2002 ACM Conference on Human Factors in Computer Systems*, 2002, p. 8.

[8] M. Ge, M. Elahi, I. Fernaández-Tobías, F. Ricci, and D. Massimo, "Using tags and latent factors in a food recommender system," in *Proceedings of the 5th International Conference on Digital Health 2015*. ACM, 2015, pp. 105–112.

[9] D. Massimo, M. Elahi, M. Ge, and F. Ricci, "Item contents good, user tags better: Empirical evaluation of a food recommender system," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization.* ACM, 2017, pp. 373–374.

[10] J. Lin, H. Pu, Y. Li, and J. Lian, "Intelligent recommendation system for course selection in smart education," *Procedia Computer Science*, vol. 129, pp. 449–453, 2018.

[11] M.-H. Hsu, "A personalized english learning recommender system for esl students," *Expert Systems with Applications*, vol. 34, no. 1, pp. 683–688, 2008.

[12] S. T. Al-Otaibi and M. Ykhlef, "A survey of job recommender systems," *International Journal of Physical Sciences*, vol. 7, no. 29, pp. 5127–5142, 2012.

[13] P. Cremonesi, C. Francalanci, A. Poli, R. Pagano, L. Mazzoni, A. Maggioni, and M. Elahi, "Social network based short-term stock trading system," *arXiv preprint arXiv:1801.05295*, 2018.

[14] (2019) Mercer quality of living survey. [Online]. Available: https://en.wikipedia.org/wiki/Mercer_Quality_of_Living_Survey

[15] (2019) Quality of living city ranking. [Online]. Available: https://www.usnews.com/news/best-countries/quality-of-life-rankings

[16] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge & Data Engineering*, no. 6, pp. 734–749, 2005.

[17] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–59, 1997.

[18] R. Burke, "Hybrid recommender systems: Survey and experiments," *User modeling and user-adapted interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[19] F. Meyer, "Recommender systems in industrial contexts," *arXiv preprint arXiv:1203.4487*, 2012.

[20] M. Lerato, O. A. Esan, A.-D. Ebunoluwa, S. Ngwira, and T. Zuva, "A survey of recommender system feedback techniques, comparison and evaluation metrics," in *2015 International Conference on Computing, Communication and Security (ICCCS).* IEEE, 2015, pp. 1–4.

[21] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The adaptive web*. Springer, 2007, pp. 291–324.

[22] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002. [Online]. Available: ./papers/burke-umuai-ip-2002.pdf

[23] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 6, pp. 734–749, June 2005.

[24] M. Elahi, "Empirical evaluation of active learning strategies in collaborative filtering."

[25] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. in Artif. Intell.*, vol. 2009, pp. 4:2–4:2, Jan. 2009. [Online]. Available: http://dx.doi.org/10.1155/2009/421425

[26] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. Kantor, Eds. Springer Verlag, 2011, pp. 1–35.

[27] Y. Koren and R. Bell, "Advances in collaborative filtering," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. Kantor, Eds. Springer Verlag, 2011, pp. 145–186.

[28] C. Desrosiers and G. Karypis, "A comprehensive survey of neighborhood-based recommendation methods," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer, 2011, pp. 107–144.

[29] M. Balabanović and Y. Shoham, "Fab: Content-based, collaborative recommendation," *Commun. ACM*, vol. 40, no. 3, pp. 66–72, Mar. 1997. [Online]. Available: http://doi.acm.org/10.1145/245108.245124

[30] M. J. Pazzani and D. Billsus, "The adaptive web," P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Heidelberg: Springer-Verlag, 2007, ch. Content-based Recommendation Systems, pp. 325–341. [Online]. Available: http://dl.acm.org/citation.cfm?id=1768197.1768209

[31] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," in *Proceedings of the ACM SIGIR*

'99 Workshop on Recommender Systems: Algorithms and Evaluation. Berkeley, California: ACM, 1999.

[32] Q. Li and B. M. Kim, "An approach for combining content-based and collaborative filters," in *Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages - Volume 11*, ser. AsianIR '03. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003, pp. 17–24. [Online]. Available: http://dx.doi.org/10.3115/1118935.1118938

[33] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM conference on Computer supported cooperative work.* ACM, 2000, pp. 241–250.

[34] R. Chen, Q. Hua, Y.-S. Chang, B. Wang, L. Zhang, and X. Kong, "A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks," *IEEE Access*, vol. 6, pp. 64 301–64 320, 2018.

[35] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, 2004.

[36] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl *et al.*, "Item-based collaborative filtering recommendation algorithms." *Www*, vol. 1, pp. 285–295, 2001.

[37] M. Nasery, M. Elahi, and P. Cremonesi, "Polimovie: a feature-based dataset for recommender systems," in *ACM RecSys workshop on crowdsourcing and human computation for recommender systems (CrawdRec)*, vol. 3, 2015, pp. 25–30.

[38] P. Lops, M. De Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender systems handbook.* Springer, 2011, pp. 73–105.

[39] F. Bakhshandegan Moghaddam and M. Elahi, "Cold start solutions for recommendation systems," in *Big Data Recommender Systems, Recent Trends and Advances.* IET, 2019.

[40] M. Elahi, M. Braunhofer, T. Gurbanov, and F. Ricci, "User preference elicitation, rating sparsity and cold start," in *Collaborative Recommendations: Algorithms, Practical Challenges and Applications.* World Scientific, 2019.

[41] G. Shani and A. Gunawardana, "Evaluating recommendation systems," in *Recommender systems handbook.* Springer, 2011, pp. 257–297.

[42] D. C. Funder, *The personality puzzle.* WW Norton & Co, 1997.

[43] I. Cantador, I. Fernández-Tobías, and A. Bellogín, "Relating personality types with user preferences in multiple entertainment domains," in *CEUR workshop proceedings.* Shlomo Berkovsky, 2013.

[44] R. Hu and P. Pu, "A study on user perception of personality-based recommender systems," in *International Conference on User Modeling, Adaptation, and Personalization.* Springer, 2010, pp. 291–302.

[45] I. Fernández-Tobías, M. Braunhofer, M. Elahi, F. Ricci, and I. Cantador, "Alleviating the new user problem in collaborative filtering by exploiting personality information," *User Modeling and User-Adapted Interaction*, vol. 26, no. 2-3, pp. 221–255, 2016.

[46] P. J. Rentfrow and S. D. Gosling, "The do re mi's of everyday life: the structure and personality correlates of music preferences." *Journal of personality and social psychology*, vol. 84, no. 6, p. 1236, 2003.

[47] M. Elahi, M. Braunhofer, F. Ricci, and M. Tkalcic, "Personality-based active learning for collaborative filtering recommender systems," in *Congress of the Italian Association for Artificial Intelligence.* Springer, 2013, pp. 360–371.

[48] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr, "A very brief measure of the big-five personality domains," *Journal of Research in personality*, vol. 37, no. 6, pp. 504–528, 2003.

[49] P. T. Costa and R. R. MacCrae, *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual.* Psychological Assessment Resources, Incorporated, 1992.

[50] V. Benet-Martinez and O. P. John, "Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the big five in spanish and english." *Journal of personality and social psychology*, vol. 75, no. 3, p. 729, 1998.

[51] O. P. John, S. Srivastava *et al.*, "The big five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of personality: Theory and research*, vol. 2, no. 1999, pp. 102–138, 1999.

[52] G. González, B. López, and J. L. De La Rosa, "The emotional factor: An innovative approach to user modelling for recommender systems," in *Workshop on Recommendation and Personalization in e-Commerce*, 2002, pp. 90–99.

[53] M. Braunhofer, M. Elahi, and F. Ricci, "User personality and the new user problem in a context-aware point of interest recommender system," in *Information and Communication Technologies in Tourism 2015*. Springer, 2015, pp. 537–549.

[54] "Master thesis source code," https://github.com/ayoubGL/Master_thesis_Country_RS.

[55] R. D. Peng and E. Matsui, *The Art of Data Science: A guide for anyone who works with Data*. Skybrude consulting LLC, 2016.

[56] M. McAuliffe and M. Ruhs, "World migration report 2018," *Geneva: International Organization for Migration*, 2017.

[57] "Data preprocessing source code :," https://github.com/ayoubGL/Countries-Rating.git.

[58] C. Cummings, J. Pacitto, D. Lauro, and M. Foresti, "Why people move: understanding the drivers and trends of migration to europe," *London: Overseas Development Institute*, 2015.

[59] F. Willekens, D. Massey, J. Raymer, and C. Beauchemin, "International migration under the microscope," *Science*, vol. 352, no. 6288, pp. 897–899, 2016.

[60] R. Verwiebe, S. Mau, N. Seidel, and T. Kathmann, "Skilled german migrants and their motives for migration within europe," *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale*, vol. 11, no. 3, pp. 273–293, 2010.

[61] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.

[62] "Recommender System algorithms implemented in Surprise lib:," https://surprise.readthedocs.io/en/stable/prediction_algorithms_package.html,.

[63] N. Hug, "Surprise, a python library for recommender systems," *URL: http://surpriselib. com*, 2017.

[64] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, no. 8, pp. 30–37, 2009.

[65] R. M. Bell and Y. Koren, "Lessons from the netflix prize challenge." *SiGKDD Explorations*, vol. 9, no. 2, pp. 75–79, 2007.

[66] C.-C. Ma, "A guide to singular value decomposition for collaborative filtering," *Computer (Long Beach, CA)*, vol. 2008, pp. 1–14, 2008.

[67] S. Funk, "Netflix update: Try this at home (december 2006)," *URL http://sifter. org/˜ simon/journal/20061211. html*, 2006.

[68] L. Prechelt, "Early stopping-but when?" in *Neural Networks: Tricks of the trade.* Springer, 1998, pp. 55–69.

[69] "Regularization in Machine Learning:," https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a,.

[70] Y. Koren, "Factor in the neighbors: Scalable and accurate collaborative filtering," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 4, no. 1, p. 1, 2010.

[71] "Collaborative Filtering(co-clustering) :," https://datasciencemadesimpler.wordpress.com/tag/co-clustering/.

[72] T. George and S. Merugu, "A scalable collaborative filtering framework based on co-clustering," in *Fifth IEEE International Conference on Data Mining (ICDM'05).* IEEE, 2005, pp. 4–pp.

[73] M. Braunhofer, M. Elahi, and F. Ricci, "Techniques for cold-starting context-aware mobile recommender systems for tourism," *Intelligenza Artificiale*, vol. 8, no. 2, pp. 129–143, 2014.

[74] "Django documentation :," https://docs.djangoproject.com/en/2.2/,.

[75] "postgresql documentation :," https://www.postgresql.org/,.

[76] "emigration factors dataset :," https://bit.ly/2m7QIqL,.

[77] N. Rubens, M. Elahi, M. Sugiyama, and D. Kaplan, "Active learning in recommender systems," in *Recommender systems handbook.* Springer, 2015, pp. 809–846.

[78] M. Elahi, F. Ricci, and N. Rubens, "A survey of active learning in collaborative filtering recommender systems," *Computer Science Review*, vol. 20, pp. 29–50, 2016.

[79] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[80] M. Braunhofer, M. Elahi, and F. Ricci, "Usability assessment of a context-aware and personality-based mobile recommender system," in *International conference on electronic commerce and web technologies*. Springer, 2014, pp. 77–88.

[81] M. D. Ekstrand, F. M. Harper, M. C. Willemsen, and J. A. Konstan, "User perception of differences in recommender algorithms," in *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 2014, pp. 161–168.

[82] "Measuring SUS :," https://measuringu.com/sus//,.

[83] "How to measure product usability with the system usability scale (sus) score :," https://uxplanet.org/how-to-measure-product-usability-with-the-system-usability-scale-sus-score, 2018.Accessedon2019-02-17.,.

[84] "Country RSs analysis :," https://github.com/ayoubGL/Country_RS_analysis.git,.