Personalisation for (public) media Assignment 1

MSc Applied Data Science

Utrecht University

ABC iview shows recommender system

Author:
Sander Engelberts
1422138
s.engelberts@students.uu.nl

March 15, 2022

ABC iview shows recommender system

Sander Engelberts[a]

[a]*MSc Applied Data Science, Utrecht University*

Abstract

The Australian Broadcasting Corporation (ABC) has to balance their public values with values of other stakeholders like their show viewers. To tackle that, this research creates a recommender system prototype for ABC iview shows where users have the autonomy to weigh personalization against content- and representation diversity. Also, the reason why certain recommendations are displayed is justified in the interface to increase trust in the platform.

Keywords

Recommender system, ABC iview, ABC, public broadcaster, streaming service, stakeholders, value sensitive design, public values, content diversity, representation diversity, justification, autonomy

1. Introduction

Recommender systems (RS) can aid in filtering and ranking relevant content for users from a vast amount of information like retail products, restaurants, and movies [1]. These recommendations can among others be determined based on historical data of similar users (collaborative-filtering RS), item information (content-based RS), or demographic information (demographic-based RS) [1]. This research focuses on building a content-based recommender system prototype for the *Australian Broadcasting Corporation (ABC)* streaming service *iview*, which includes television shows, live performances and movies (hereafter combined referred to as shows) [2]. Such a public broadcaster needs to serve a public mission, while also balancing the values of other stakeholders [3].

To name some of these stakeholders: ABC themselves, their viewers and other people living in Australia, content creators, and the Australian government who are their main financial funder [4]. Here is focused on the public values of the ABC, which are influenced by the government, and the ones of its users to try to balance their potential value tensions.

In short, the ABC has as public mission to be an independent and trustworthy platform which shares high-quality stories that reflect the Australian society, culture, and identity for the purpose of education, entertainment, and journalism [5, 6]. To elaborate, they want to represent the diversity among Australian communities, for example cultural, age, and socio-economic diversity. Also, they want to stay relevant by innovating, be sustainable, and be accessible. Lastly, they want to deliver personalization on iview and ensure cybersafety [5, 6].

Below a selection of such (public) values is described, which will be taken into account while designing the recommender system:

- Personalization refers to recommending items that match user interests and preferences based on for example their demographics or previous watch history [1, 7]. Yet, this can

Personalisation for (public) media, March 15, 2022
1422138 (S. Engelberts)

lead to 'filter bubbles' with limited perspectives [8].

- Diversity is an often contrasting public value to personalization, which is split up into two types:
  - Content diversity can be defined as exposing the user to items that are different from their watch history and contain diverse viewpoints in order to facilitate the public debate, reduce filter bubbles, and give more choices [9, 10]. Different users prefer different levels of content diversity, so may be given autonomy [9].
  - Representation diversity can be defined as presenting a diverse range of representations of minority/oppressed identities to reduce biases [11, 12].
- Transparency refers to explaining the inner logic of the algorithm to the user, whereas justification only describes why they receive certain recommendations instead of how, which may be preferred for simplicity [10]. When these are higher, the users trust the system more [13], which is important to the ABC.
- Autonomy gives the user influence over their presented recommendations by for example changing preference settings [10, 12].

To check how these (public) values align with the values of potential users, I set out a short survey[1]. From this could be inferred that respondents want at least a bit of autonomy; they differ in the extent in which they prefer more personalized or more content diverse recommendations; they prefer content diversity over representation diversity to some degree. Because of these results, a certain level of autonomy for selecting preferences with regards to recommendations was implemented, while taking into account that for all users at least a minimum level of ABC's public values is enforced.

2. Data

The used data consists of 1538 .html files that the teachers of the course Personalisation for (public) media at Utrecht University (2022) scraped from the ABC iview website. Because shows can cover multiple genres and the files were ordered in subfolders based on genres, duplicate files needed to be removed by merging genre labels. Next to retrieving the genre information from the subfolder names, the show title, description, webpage url, and image url were gathered from the .html page using `BeautifulSoup`.

With regards to further data exploration, Figure 1 shows that the descriptions are all relatively short, which impacts how accurately the content of shows can be inferred. Due to this, shows with descriptions shorter than 20 words were removed, which resulted in a dataset of 1047 shows (92% of the 1136 unique shows).

Moreover, in Table 1 can be seen that some genres are (much) more represented in the dataset than other genres. However, the description length statistics do not vary much between the genres, which reduces the difference between genres in how accurately the content of a show can be inferred.

[1] For the questions and results see Appendix A.

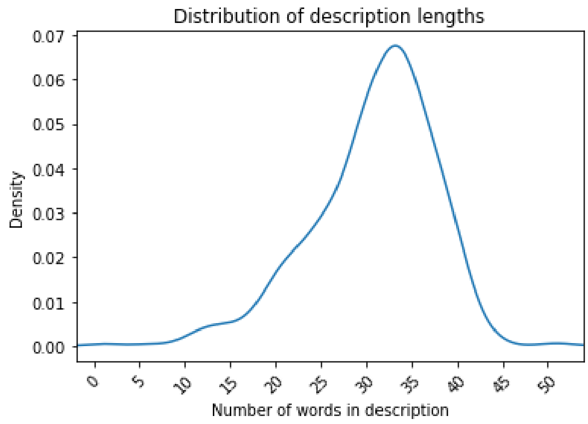


Figure 1: Distribution of description lengths of ABC iview shows, with the statistics: minimum 1 word, maximum 52 words, mean 31 words, and median 32 words.

Genre	**Number of shows**	**Frequency of genre**
Drama	231	16.2%
Documentary	197	13.8%
Education	163	11.4%
Kids	153	10.7%
Family	145	10.2%
Movies	137	9.6%
Comedy	121	8.5%
Arts and culture	109	7.6%
News and current affairs	90	6.3%
Lifestyle	48	3.4%
Panel and discussion	33	2.3%

Table 1
Distribution of shows over genres. For conciseness here the shows are contained in all the genre labels they belong to, giving a total of 1427 shows of the 1047 unique shows with descriptions longer than 20 words.

3. Methods

After obtaining, cleaning and exploring the dataset, tokenized and vectorized data was clustered into shows with similar content, and 'distances' between shows were computed to be able to rank recommendations with regards to public values. Here, the show descriptions were used to get an indication about the content of shows, and the genres were used for additional information about the type of movie.

The created code, show information and recommendation datasets, and the final streamlit interface app are made available via GitHub.

3.1. Pre-processing

The natural language processing tool `spaCy` was used to tokenize and filter the description text with appended genre label(s) of each show. With regards to filtering, punctuation and stop words were removed, tokens were lower-cased and lemmatized, and only the following word types were kept: proper nouns, adjectives, verbs, and nouns. This way non-informative words regarding the content of shows were removed.

After, the tokens were vectorized using `TfidfVectorizer` of `scikit-learn` to numerically represent the texts [14]. Tokens occurring in more than 90% of the show texts were also removed (i.e. $max\_df = 0.9$) as these are non-indicative of distinguishing the shows, and no minimum word occurrence was applied (i.e. $min\_df = 0$), because only a few shows may contain words about specific minority/ oppressed groups.

3.2. Clustering

Next, a K-Means clustering model was trained to extract groups of shows with similar text characteristics. Because there are multiple genres assigned to some shows and within genres the content of shows can differ a lot, the genre labels were not directly used to cluster the shows, but instead K-Means with 15 clusters was applied to extract the content cluster a shows fits best in. There are 11 distinct genres in the dataset, from which 47 combinations are used to label shows, and of these only 17 combinations contain more than 10 shows. It is thus reasonable to create between 11 and 17 clusters of shows to balance the level of content detail such a cluster is based on, as an 'elbow' technique also indicated [15].

3.3. Ranking recommendations

Using these content clusters and the numerical tf-idf vector representations of show texts, shows were ranked for different public values to make a top 10 selection of recommendations for each show. These top 10 recommendations for each public value were then combined to form a top 10 of overall recommendations (see Section 4).

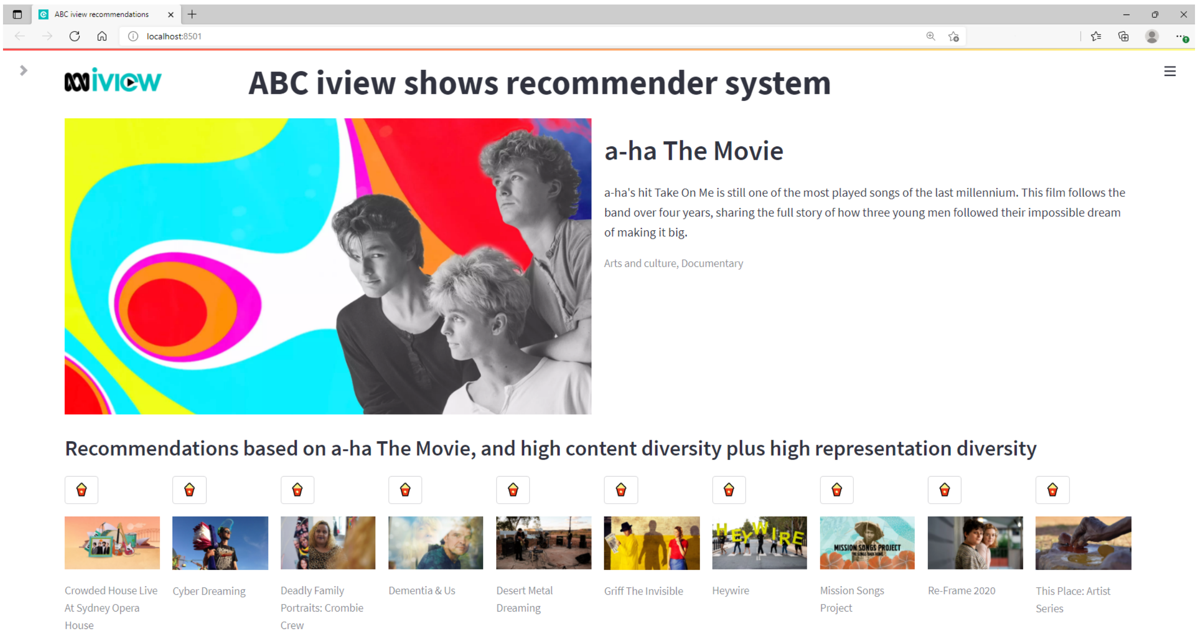
This ranking is done by calculating the cosine distance between the vector representations of show texts [16]. Similar shows from the same cluster are said to give personalized recommendations because these are similar in content and type of show, whereas shows from a different cluster broaden the user's content diversity.

Next to calculating the most similar shows within and outside the same cluster, also the distance to shows that represent minority/ oppressed groups was calculated to be able to enhance representation diversity. For this, a collection of 73 shows that represent these communities was generated by checking if a word describing such groups (from [17, 18, 19, 20]) was included in their tokens.

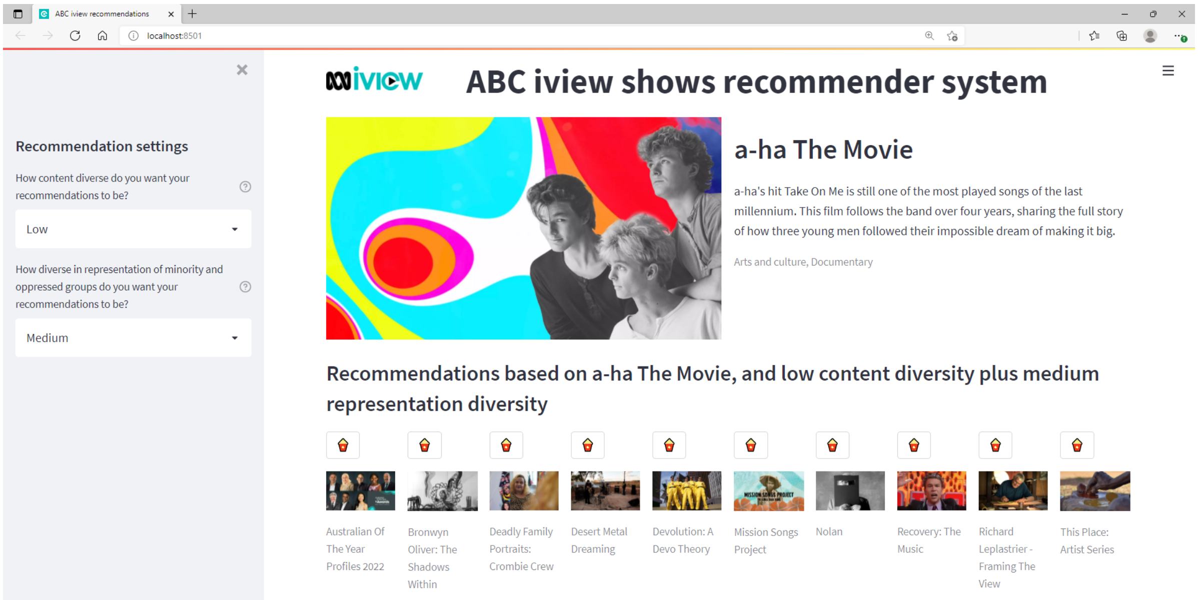
4. Interface design

The interface of the ABC iview recommender system displays, as can be seen in Figure 2, the currently selected show with its image, title, description, and genre label(s). Underneath are its top 10 recommended shows based on the content of this selected one.

To give users the autonomy of selecting how much they value content- and representation diversity, settings were created in a side bar to change each of these between 'Low', 'Medium' (default), and 'High' (see Figure 2b). This influences in turn how many of the displayed top 10 recommended shows are from which list of most similar shows: personalized (within cluster) shows, content diverse (outside cluster) shows, or representation diverse shows. Then the order of this combination of shows is randomized to not preference one value over the other while displaying. When a user is not interested in receiving diverse shows (i.e. sets diversity to



(a) Interface showing recommendations for *a-ha The Movie* with high content and representation diversity settings.



(b) Interface with the diversity settings sidebar expanded, showing recommendations for *a-ha The Movie* with low content and medium representation diversity settings.

Figure 2: ABC iview recommender system interface prototype.

'Low'), then they still get 1 show for each diversity type to enforce a minimum level of ABC's public values.

Furthermore, the settings for types of diversity also have information buttons to get a further explanation about what this entails, to enhance justification. Because users may not check the settings, the header above the recommendations also shows what settings the recommendations are based on and that these are generated for the specific show that is currently viewed.

5. Conclusion and discussion

This research shows that it is possible to determine show recommendations that are on the one hand personalized by being similar in content to a selected show, and on the other hand contains a larger content diversity. Additionally, similarity to shows representing minority/oppressed groups can be determined, and users can be given autonomy to choose how they want these public values to be balanced. Moreover, justification can be increased by simply adding feedback why certain recommendations are shown.

There are some shortcomings to this prototype however. Firstly, the data contained only short descriptions from which it may not always be possible to accurately infer the contents of a show. This results in less accurate clustering of shows and hence less accurate recommendations.

Similarly, show descriptions do not always explicitly refer to the communities they represent, so some shows which actually do enhance representation diversity may be falsely missed. On the contrast, a word like 'black' may not refer to a person's skin color in a certain description, giving incorrect labeling as well. This could be tackled by manual labeling shows for representation diversity.

Furthermore, even though tf-idf is a simple and effective method that does not require too much computing power, it does not capture the context of words like word embeddings would. A model like BERT may thus be more accurate at inferring the content of a show and hence give better recommendations.

Additionally, to be more in line with the sustainability value of ABC, and to make the model scalable to a bigger dataset of shows, the distance computations should be optimized. Currently it is quick to calculate the pairwise similarity between all shows, but it would be better to only calculate this towards its nearest neighbors within its own cluster and the closest other clusters. Then should also be taken into account that shows that are near the border of a cluster may be more similar to shows in a neighboring cluster than to some of the top 10 within cluster recommended shows.

To end, public broadcasters like ABC have an additional challenge of incorporating public values into their recommender systems and balance these with the values of other stakeholders. So it is valuable to continue exploring different types of recommender systems and operationalisations of these values to be able to gain more valuable recommendations that broaden someones horizon while still being personalized to their interests.

References

[1] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, R. Kashef, Recommendation systems: Algorithms, challenges, metrics, and business opportunities, Applied Sciences 10 (2020) 7748. doi:10.3390/app10217748.

[2] Australian Broadcasting Corporation, ABC iview login to watch, 2022. https://about.abc.net.au/press-releases/%e2%80%afabc%e2%80%afiview%e2%80%aflogin-to-watch/.

[3] A. Smets, N. Walravens, P. Ballon, Designing recommender systems for the common good, in: Adjunct Publication of the 28th ACM Conference on User Modeling, Adap-

tation and Personalization, UMAP '20 Adjunct, Association for Computing Machinery, New York, NY, USA, 2020, p. 276–278. URL: https://doi.org/10.1145/3386392.3399570. doi:10.1145/3386392.3399570.

[4] R. Jolly, The ABC: an overview, 2014. https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1415/ABCoverview#_Toc395086104.

[5] Australian Broadcasting Corporation, ABC five-year plan 2020-2025, 2020. https://about.abc.net.au/wp-content/uploads/2020/07/ABC-Five-Year-Plan-FINAL-Updated.pdf.

[6] Australian Broadcasting Corporation, Corporate responsibility, n.d. https://about.abc.net.au/how-the-abc-is-run/what-guides-us/corporate-responsibility/.

[7] D. Shin, B. Zhong, F. A. Biocca, Beyond user experience: What constitutes algorithmic experiences?, International Journal of Information Management 52 (2020) 102061. URL: https://www.sciencedirect.com/science/article/pii/S0268401219314161. doi:https://doi.org/10.1016/j.ijinfomgt.2019.102061.

[8] A. Bruns, Filter bubble, Internet Policy Review 8 (2019). URL: https://policyreview.info/concepts/filter-bubble. doi:https://doi.org/10.14763/2019.4.1426.

[9] N. Helberger, K. Karppinen, L. D'Acunto, Exposure diversity as a design principle for recommender systems, Information, Communication & Society 21 (2018) 191–207. URL: https://doi.org/10.1080/1369118X.2016.1271900. doi:10.1080/1369118X.2016.1271900. arXiv:https://doi.org/10.1080/1369118X.2016.1271900.

[10] C. He, D. Parra, K. Verbert, Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities, Expert Systems with Applications 56 (2016) 9–27. URL: https://www.sciencedirect.com/science/article/pii/S0957417416300367. doi:https://doi.org/10.1016/j.eswa.2016.02.013.

[11] L. Schelenz, Diversity-Aware Recommendations for Social Justice? Exploring User Diversity and Fairness in Recommender Systems, Association for Computing Machinery, New York, NY, USA, 2021, p. 404–410. URL: https://doi.org/10.1145/3450614.3463293.

[12] B. Friedman, P. Kahn, A. Borning, Value sensitive design: Theory and methods (2003).

[13] P. Pu, L. Chen, R. Hu, A user-centric evaluation framework for recommender systems, in: Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 157–164. URL: https://doi.org/10.1145/2043932.2043962. doi:10.1145/2043932.2043962.

[14] J. Silge, D. Robinson, Analyzing word and document frequency: tf-idf, 1 ed., O'Reilly Media, Inc., 2017. URL: https://www.tidytextmining.com/tfidf.html.

[15] N. K. Ersi, K-means elbow method code for Python, 2020. https://predictivehacks.com/k-means-elbow-method-code-for-python/.

[16] J. Möller, D. Trilling, N. Helberger, B. van Es, Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity, Information, Communication & Society 21 (2018) 959–977. URL: https://doi.org/10.1080/1369118X.2018.1444076. doi:10.1080/1369118X.2018.1444076. arXiv:https://doi.org/10.1080/1369118X.2018.1444076.

[17] N. Ghanea-Hercock, Ethnic and religious groups in the Islamic Republic of Iran, 2004. https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww2.ohchr.org%2Fenglish%2Fissues%2Fminorities%2Fdocs%2FWP2.doc&wdOrigin=BROWSELINK.

[18] T. R. Conerly, K. Holmes, A. L. Tamang, Race and Ethnicity, 3 ed., OpenStax, 2021. URL: https://openstax.org/books/introduction-sociology-3e/pages/11-1-racial-ethnic-and-minority-groups.

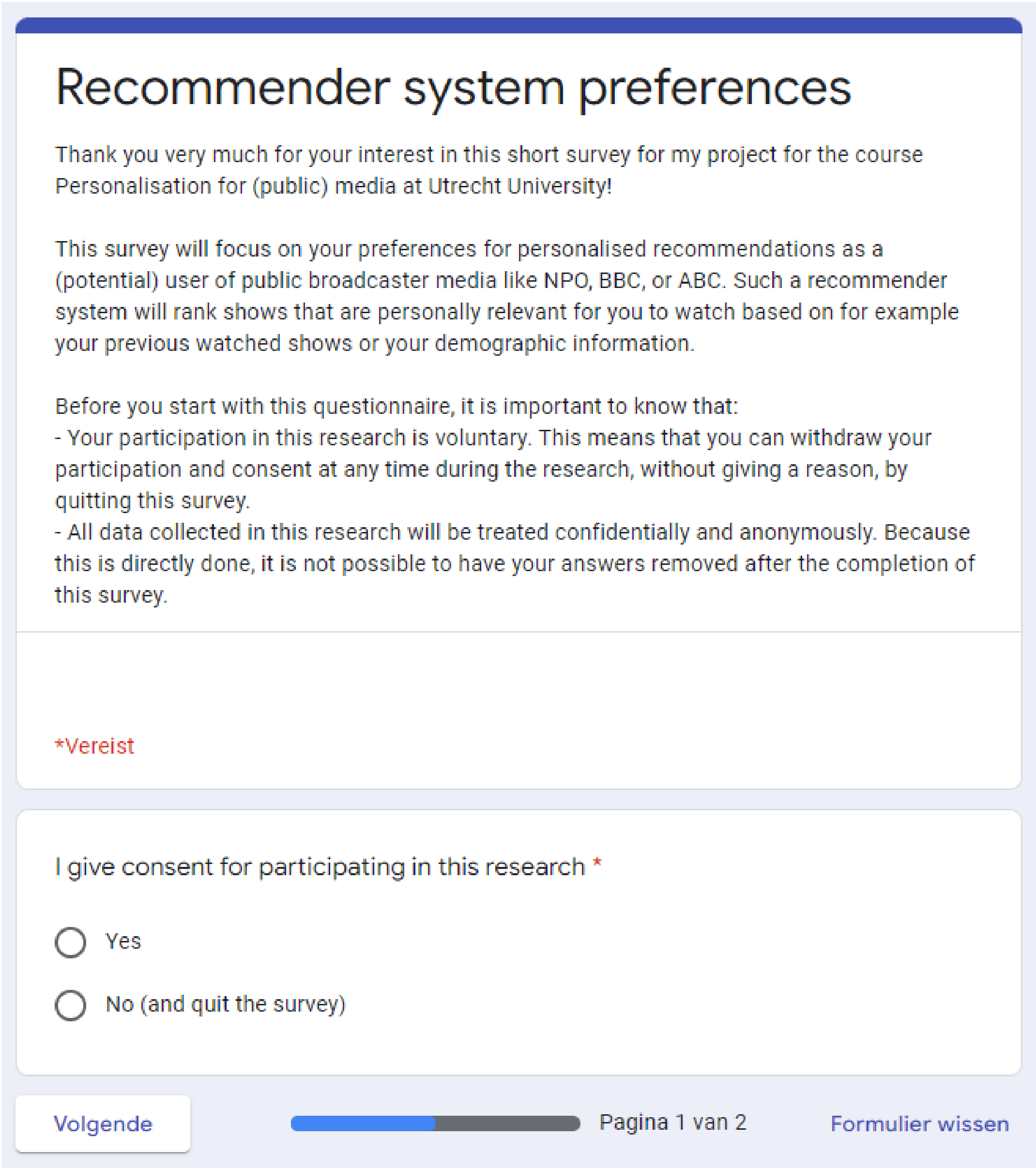
[19] S. S. Thomas, T. Andrews, R. Varina, 45 gender- and sexuality-related terms you need to know, 2021. https://www.cosmopolitan.com/sex-love/a24442815/sexual-orientation-list-definitions/.

[20] O. Yiftachel, Ethnic conflict, in: R. Kitchin, N. Thrift (Eds.), International Encyclopedia of Human Geography, Elsevier, Oxford, 2009, pp. 601–607. URL: https://www.sciencedirect.com/science/article/pii/B9780080449104007756. doi:https://doi.org/10.1016/B978-008044910-4.00775-6.

A. Survey

Figure 3 shows the conducted *Google Forms* survey[2] and its results. Note that these results are not representative for the ABC, because the contacted respondents were Applied Data Science students, so are (likely) not Australian citizens. Moreover, only 13 responses were collected, which can not represent Australia's full diversity of people.

[2] Click here for the online version.



Recommender system preferences

Thank you very much for your interest in this short survey for my project for the course Personalisation for (public) media at Utrecht University!

This survey will focus on your preferences for personalised recommendations as a (potential) user of public broadcaster media like NPO, BBC, or ABC. Such a recommender system will rank shows that are personally relevant for you to watch based on for example your previous watched shows or your demographic information.

Before you start with this questionnaire, it is important to know that:
- Your participation in this research is voluntary. This means that you can withdraw your participation and consent at any time during the research, without giving a reason, by quitting this survey.
- All data collected in this research will be treated confidentially and anonymously. Because this is directly done, it is not possible to have your answers removed after the completion of this survey.

*Vereist

I give consent for participating in this research *

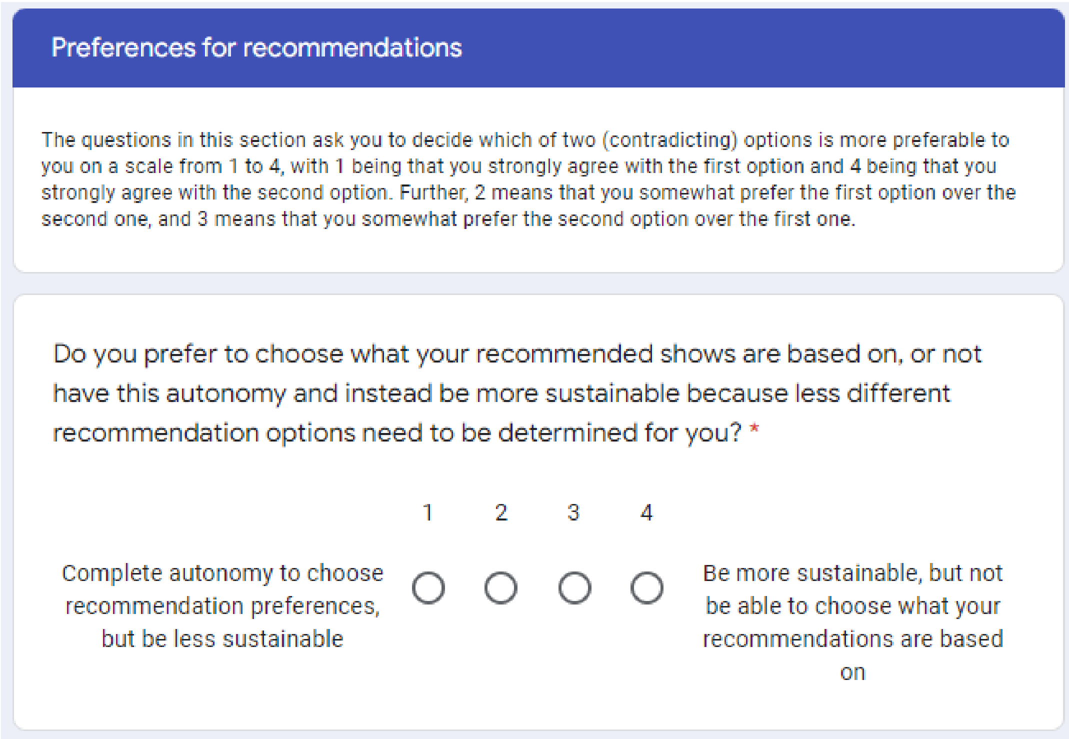
- Yes
- No (and quit the survey)

Volgende

Pagina 1 van 2

Formulier wissen

(a) Introduction and consent page.



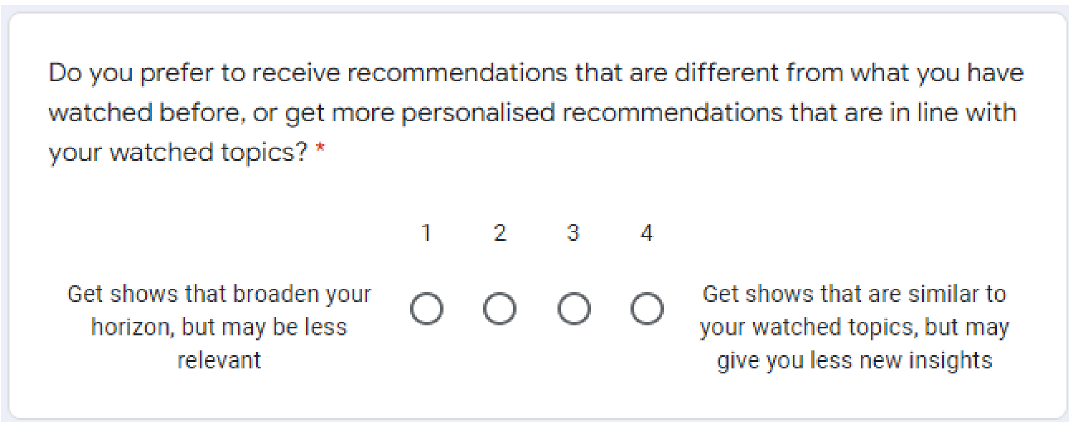
Preferences for recommendations

The questions in this section ask you to decide which of two (contradicting) options is more preferable to you on a scale from 1 to 4, with 1 being that you strongly agree with the first option and 4 being that you strongly agree with the second option. Further, 2 means that you somewhat prefer the first option over the second one, and 3 means that you somewhat prefer the second option over the first one.

Do you prefer to choose what your recommended shows are based on, or not have this autonomy and instead be more sustainable because less different recommendation options need to be determined for you? *

	1	2	3	4	
Complete autonomy to choose recommendation preferences, but be less sustainable	○	○	○	○	Be more sustainable, but not be able to choose what your recommendations are based on

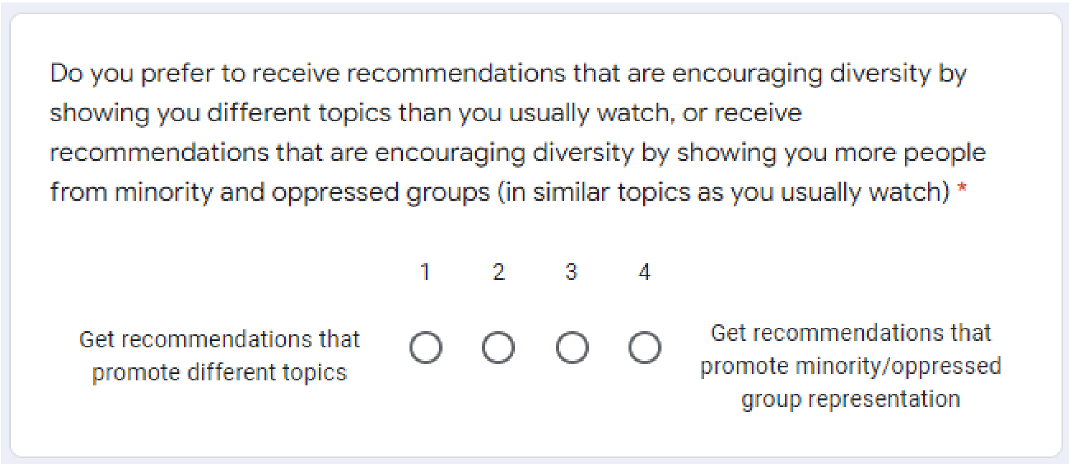
(b) Autonomy versus sustainability. The number and percentage of respondents that selected each option from 1 to 4: 3 (23.1%), 4 (30.8%), 6 (46.2%), 0 (0.0%).



Do you prefer to receive recommendations that are different from what you have watched before, or get more personalised recommendations that are in line with your watched topics? *

	1	2	3	4	
Get shows that broaden your horizon, but may be less relevant	○	○	○	○	Get shows that are similar to your watched topics, but may give you less new insights

(c) Personalization versus content diversity. The number and percentage of respondents that selected each option from 1 to 4: 1 (7.7%), 5 (38.5%), 3 (23.1%), 4 (30.8%).



(d) Content diversity versus representation diversity. The number and percentage of respondents that selected each option from 1 to 4: 3 (23.1%), 5 (38.5%), 3 (23.1%), 2 (15.4%).

Figure 3: Conducted survey with its results from 13 respondents.