**RESULTS OVERVIEW**

Previous results may be not significant in statistical terms, since just a few users (20) are shared among trainset and testset. Moreover, 1's and 0's represent an highly imbalanced problem and this fact explains the very high accuracy value.

For this reason, a two-step roadmap is proposed to further develop this preliminary analysis and deploy the reccomendation engine.

**NEXT STEPS**

Once developed a promising reccomendation engine, there are two main steps to assess its performance

1. Make an ensemble of reccomendation engine algorithms (SVD, Collaborative Filtering, Content Based, Ranking based...) to cover all the possible situation (new user to reccomend, new articles to be reccomended...). The code deployed should follow the guidelines learning in the Part 2 of Data Science Nanodegree Program about documentation and modularity.
2. Develop an Experiment to understand what is the best way to assess real-life performance of the model

**EXPERIMENTAL DESIGN**

**A/B Test** approach is a solid procedure to understand if the model brings real benefits.

* **Cookie-based** A/B splitting allows to randomly assign a user to one of the two groups: one group uses the old homepage featuring suggestions for articles according to the "old" reccomendation engine. The second group interacts with the same homepage featuring the new reccomendation engine.
* **Funnel** Independently from the group which each user is assigned to, we expect a user entering the homepage to follow this steps: 1) visit the homepage 2) scroll the homepage looking for suggested articles 3) click on a particular article. In this scenario, we are not considering a user that wants to click on the search bar and input some keywords to find specific articles.
* **Invariant Metric** If the assignment of users to the two homepages is truly random, we expect the number of cookies hitting the two homepages not to be significally different. A proper statistical test must be performed to assess this point
* **Evaluation Metric** A possible choice could be the ration #clicks-on-reccomendations / #cookies. If our new reccomendation system is actually better than the previous one, we expect an higher number of clicks on the recommended articles made by each user. Hence, the average number of clicks on reccomendation per user in the experiment group must be greater than the control group value. Other choices are possible, for instance #visits / #cookies, if we suppose that users of the experiment group are more willing to come back to the homepage to have more reccomendations. Another one could be #clicks-on-reccomendation / #clicks-on-searchbar, since we expect that each user now has less need to look for something particular in the searchbar of the website, since interesting articles are already suggested in the homepage. Both solutions could be good, but #clicks-on-reccomendations / #cookies seems to be the most striaghtforward idea.
* **Experiment Sizing** We need to understand which statistical test should be performed and how large the two groups must be to have, let's say, an appropriate significant level that guarantees a Type I error lower than 0.05. We need to combine this kind of analysis with the average number of visits per day to understand how long the experiment should be to meet statistical constraints. The duration of the experiment must be a reasonable amount of time for the owner of the website.
* **Validity** Evaluation metric is aligned with our experiment goal: understand whether or not the new reccomendation engine has better performance than the old one. The biggest assumption we make is that users are homogeneously distributed between control and experiment group. Compared to other scenarios, this seems to be not too difficult to obtain, because the users of a data science community website are a very particular audience, it is not such an heterogeneous group of people. In case of Netflix, many factors (age, location, sex, hour of the day...) are important to understand which kind of user is visiting the website, and the previous assumption is for sure more difficult to be proved.
* **Bias** We don't expect sampling biases to arise, for the same reasons expressed in the previous bullet point: user population visiting this website can be pretty homogeneous. However, there could be some issues with novelty bias, since in the first days after signing up a user might visit the website more often than the regular frequency. If we have some previous analitycs about this behaviour (let's say, the 3 days after sign up are always very different from the following period), we should take it into account (the experiment must be much longer than 3 days).
* **Ethics** The change under-the-hood of the homepage should be benign to each user. Moreover, no particular personal data are required, just the user-article interaction. The usual message about cookie regulamentation must appear for each user as usual.