### Validation H6 recommender system - rules and limitations

***Test system:***

100 test clients are made and, for every user, 1-3 features are randomly generated by the model. Based on these features, for every test client, the model will generate (10 to 30) pageviews (10 museum pages that can be clicked on 1 to a max. of 3 times) that contain these features. Then, it uses these pageviews for its calculations to recommend 10 museums.

This system is tested by following the rules below. Recommendations for a test client are classified as ‘good’ if they pass all rules. Then, with more test clients, we can see how well our system performs.

Next to the recommending based on pageviews, we want to enhance our model’s performance by using search queries from the search engine as features. We searched for the 15 most searched for keywords (mostly locations) and used these.

Implemented features:

- facilities (taxonomy data)

- themes (based on external source)

Features we are working on:

- Opening hours

- Museum entrance price

- Clusters of museum description

- Clusters of asset pictures

***Rules (bold numbers are being discussed):***

- If one feature is given to the system:

- feature appears in at least **70%** of recommendations (recs)

- If 2 or 3 features are given to the system:

- all features appear in at least **40%** of recs

- check the performance per feature. If all individual features pass, the rule is passed.

- At least **70%** of recs is small museum

- treshold of small/large museum is based on expected #visitors of 2019. The median number of visitors is taken as critical point, so all museums with that or less expected visitors are considered small museums.

- At least **2** exhibitions/events are in recs

- we implement this by recommending for 10 museums and then 2 museums are randomly chosen and converted to an exhibition or event of that event.

- Displayed exhibitions are available now or in future (so no past exh.)

- If search queried a specific location: at least **30%** comes from that location/province

- If search queried several locations: at least **a museum** from **all** those **locations** in rec

***Limitations/Assumptions:***

*-* The test system uses pageview clicks based on the features we give. However, it is possible that people searching museum.nl do not have clicks exclusively with those features. Still, we assume that most people’s searches are directed and this is the only way to generate useful recommendations.

- ..

***Statistics, test on many test clients:***

- Check individual features. Which perform well/bad?

- Check individual museums. Do some museums never/always appear?

- How many requested features does a museum have on average? E.g. if 3 features are given, do all recommendations contain either 1 of those features or do they contain combinations of them?

- ...