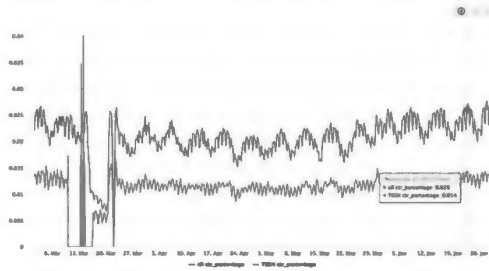

details in [post](https://b.workplace.com/notes/668607013449104), FS notebook: [N3352710](https://www.internalfb.com/intern/amp/view/?id=N3352710)) **Results**.* * We compare the NE % difference between models trained with features selected by teen data versus overall data for two scenarios (1) model trained on teen data and (2) model trained on overall data. Results show that ** for both scenarios, features selected based on teen data does not significantly improve teen NE**.* * We checked the feature tags selected by teen data and overall data and found that there are about [redacted] features that are in the top [redacted] for teens but not in top [redacted] for the overall users. However, among these [redacted] features there is only [redacted] in top [redacted] and in top [redacted] based on the feature importance run based on the teen tag.** This indicates that when feature tag contains a good amount of features (say [redacted]), the tag that is selected by teens does not differ too much from that by overall data.**.* * We didn't include in the diagram above for "external share" (due to scale), the NE regressed quite a lot (around [redacted]%) for model trained with features selected by teen data. Our guess is that this is due to the low CTR for this label, and using teen data alone could not find good features for such sparse actions.**.* * Suggestions for next half**.* * We showed that the model performance based on the tags selected by teen data and by overall data are very neutral and we showed that it is due to that the features selected based on teens data is not very different i.e. among the [redacted] different features only very few of them are important, from those based on overall data and_. We should check if this is true for other nmtl models e.g. the toprank model.* * While the ecosystem changes over time and newly are registered we should check at a regular basis if there are features left out by MB, which selects features based on overall data, but are in fact important for teens. We did this this half based on per task feature importance for teens for important tasks likes, profile taps, and reshares and currently these proposed features are included in 2023 H1 MB. #thanks Special thanks to [redacted] who spent time discussing results from these projects and giving me so many invaluable feedbacks #thanks [redacted] for constant check-ins and giving feedbacks on this work #thanks [redacted] for their work on data upsampling that facilitates the analysis #thanks [redacted] for discussions of teen related features work #thanks Israel Nir for his advice on Swiss Cheese which makes this work much more easier



Thanks Bot on June 30, 2023 at 2:16 PM

Thanks for thanking, [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], Di Liu, [REDACTED], [REDACTED], [REDACTED], and [REDACTED] can see your thanks at furl.com/thanks. Awesome! For more info about Thanks formatting, visit <https://furl.com/thanks-parsing>.

on June 30, 2023 at 2:17 PM

#sharebot Relevance & Teens Insights

Move Bot on June 30, 2023 at 2:17 PM

Alright [REDACTED], I shared this post to Relevance & Teens Insights. Here is the shared post: <https://fb.workplace.com/groups/526099235989340/permalink/607251707874092/>

on June 30, 2023 at 2:18 PM

CC BY-NC-ND 4.0 International license.

on June 30, 2023 at 2:40 PM

This is fantastic summary of H1, congrats [REDACTED]! Two things I've discussed with [REDACTED] a couple of weeks ago: 1. We might benefit for not upsampling all "non-teens" uniformly. Since you've found that teen data is more efficient than non-teen data for teens, it makes sense that among non-teens we might want to over sample those that are "teen like". In the most basic form it means - sample more 20 year olds than 60 year olds. 2. Personally (and I'll need you guys' guidance here) - I'd be interested in taking the existing model and trying to fine tune it on teen data, just to see what happens.

on July 10, 2023 at 2:57 PM

Sorry for my late reply! These are great feedbacks! I thought about (1) but hasn't started any test, I think these can be done quickly on my side. For (2) let's take it offline. brought up this idea some time ago too I think we can take a try together.