# **Evaluating The Effects of Calibrated** Popularity Bias Mitigation: A Field Study

Anastasiia Klimashevskaia 1 (anastasiia.klimashevskaia@uib.no), Mehdi Elahi 1, Dietmar Jannach 2, Lars Skjærven 3, Astrid Tessem 3, Christoph Trattner 1

Media Futures



17th ACM Conference on Recommender Systems Singapore, 18th-22nd September 2023

## Problem

<sup>3</sup> TV 2, Norway

<sup>1</sup> University of Bergen, Norway

<sup>2</sup> University of Klagenfurt, Austria

- Setting: TV 2 Norwegian national broadcaster, experiencing issues with **Popularity Bias** – a tendency of a Recommender System to excessively promote highly popular items
- The effect is amplified by the feedback loop
- Leads to reduced discovery, diversity and novelty of content, lower user engagement and interest
- Goal: Develop and evaluate a re-ranking method in an A/B test to mitigate bias, increasing user engagement and content diversity

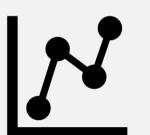
# A/B Test Design



Calibrated Popularity (CP) – re-ranking that actively accounts for historical user popularity preference, while attempting to maintain high predicted relevance



Three consecutive A/B testing periods with varying conditions and parameters (see flowchart below)



Tracking Click-Through Rates, Item Exposure and Watches

#### 19.12.22 - 19.01.23

Group B (treatment group) receiving CP-reranked recommendations with top 5% of most popular items considered as highly popular



#### 19.01.23 - 23.02.23

Item popularity threshold was adjusted to 2.5% instead of 5%



#### 1.03.23 - 29.03.23

"Randomized" version of CP, relevancy of exchanged items is not considered, tail items to be included in the recommendation are picked randomly and inserted in the list of recommendations.

### Results and Observations

We observed CTR increase for group B in Phases 1 and 3 (+1.25% and **+2.25** respectively), **CTR drop** in Phase 2 (-4.18%)

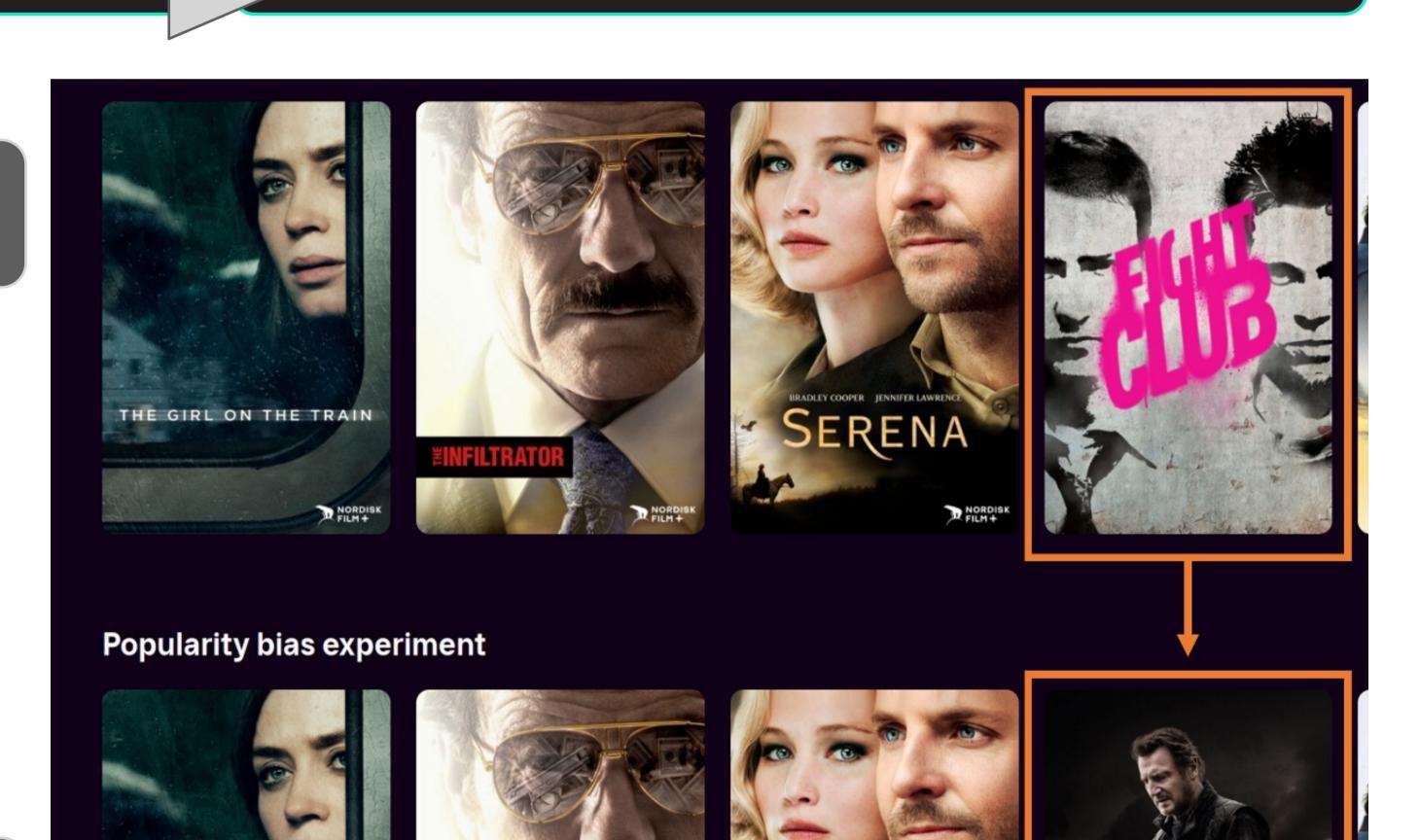
- Calibrated re-ranking can be an effective technique to provide less mainstream recommendations without negatively affecting user experience
- Thresholds and parameters used for the method need to be tuned and selected with caution
- Otherwise undesired effects can be observed
- Recommending a small amount of random items can also have positive effects on user exploration

More unique movie titles exposed to users in group A, while more unique movie titles watched by users in group B during all three phases

- This can be attributed to users in group A scrolling the recommendations more in search of novelty and serendipity
- At the same time group B requires less search and receives more novel recommendations



Read full paper



The CP Algorithm seeks to swap highly mainstream movies with less popular ones with relatively high predicted relevancy to lower the mainstreamness of recommendation

SERENA

#### **PARTNERS**

amedia Bergens A Tidende



VIMOND













#### **HOST**



#### **FUNDER**

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339)



THE MARKSMAN



**Schibsted**