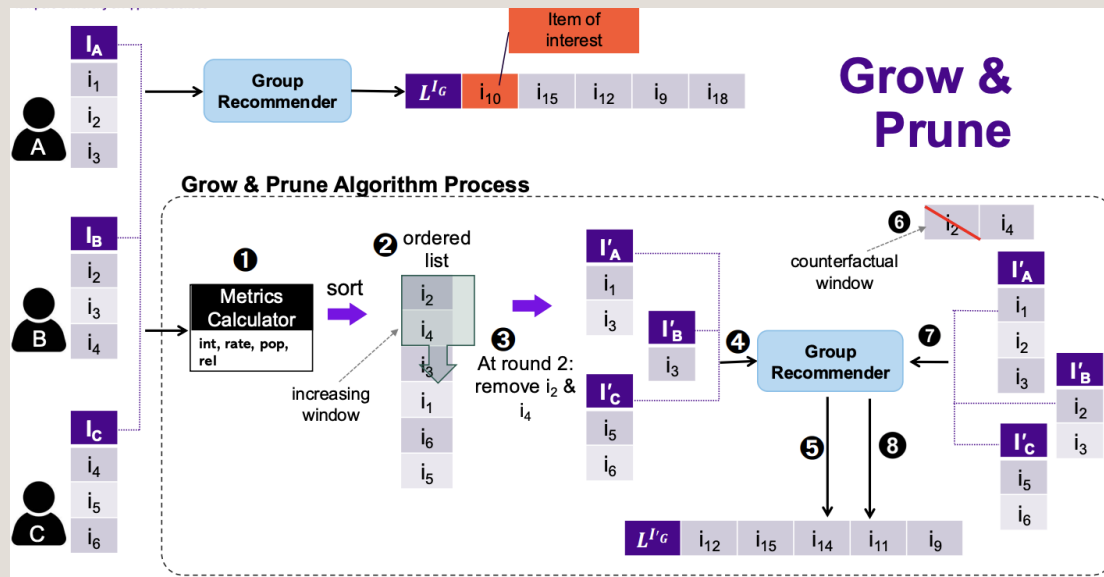


Part IV: Counterfactual Explanations for Group Recommendations



- Students: Oskari Perikangas, Xiaosi Huang
- Course: DATA.ML.360-2025-2026-1
- Date: November 28, 2025



Problem Statement: 🤔

**"If the group had NOT watched items A,
then item B would NOT be recommended"**

Challenge:

Ensure fairness - no single-user blame in group explanations

Technical Stack:

Grow & Prune Algorithm

Pareto Filtering (Multi-objective optimization)

Item-based KNN Recommender

4 metrics (recognition, rating, influence, explanatory power)

+ fairness constraint ≥ 0.67



Metrics & Functions

1. Item-Level Metrics (Slides p.12-14)

Recognition: Measures how many group members rated an item

$$rs(i, G) = \frac{|\{u \in G : i \in I_u\}|}{|G|}$$

Items known by the majority create more understandable and fairer explanations.

Average Rating: Group's average rating for an item

$$rate(i, G) = \frac{\sum_{u \in G: i \in I_u} r_{u,i}}{|\{u \in G : i \in I_u\}|}$$

where $r_{u,i}$ is the rating user u gave to item i . High-rated items have stronger influence on recommendations.

Influence: Impact on target recommendation

$$infl(i, t, G') = \frac{\sum_{u \in G'} \hat{r}_{u,t}}{|G'|}$$

where $G' = \{u \in G : i \in I_u\}$ and $\hat{r}_{u,t}$ is the predicted rating for target t . Ensures explanation items actually contribute to recommending the target.

Explanatory Power (fast estimation):

$$expwr(i, t, G) = sim(i, t) \cdot \frac{rate(i, G)}{5} \cdot rs(i, G)$$

where $sim(i, t)$ is item-item similarity from KNN model. Estimates influence without rebuilding recommender (reduces computation). Full validation performed during GROW and PRUNE phases.

2. Fairness Constraint (Slides p.16)

To prevent single-user blame, we require at least 2/3 of group members to contribute:

$$fair(G, E) = \frac{|\{u \in G : \exists i \in E, i \in I_u\}|}{|G|} \geq 0.67$$

For a 3-person group, this means ≥ 2 members must have rated items in the explanation. Ensures collective responsibility.

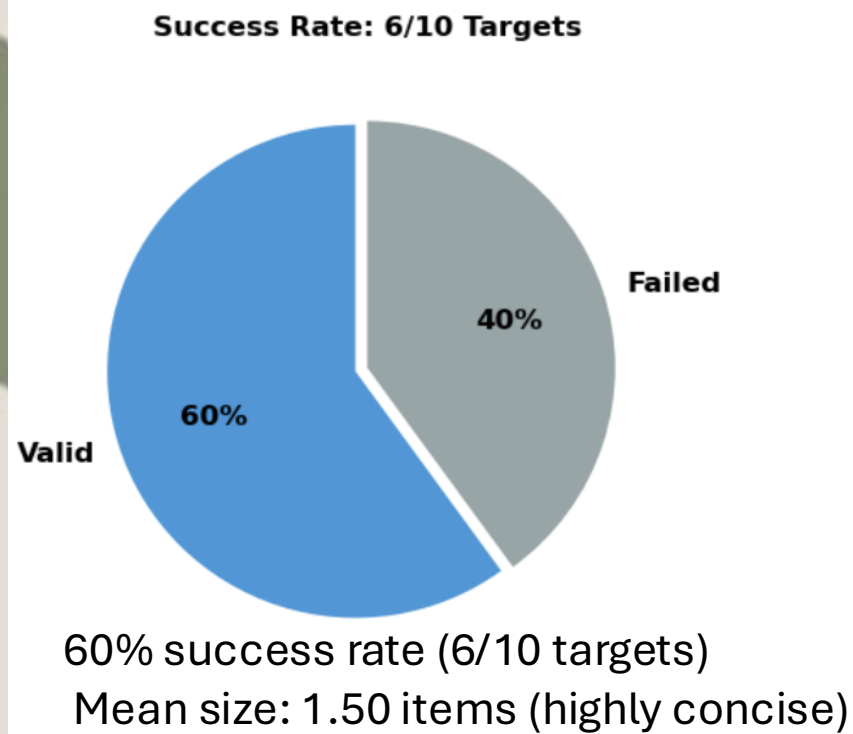
Solution Flow _ Input: Group G, Target t, Top-N list L

1. Get Candidates (~1275 movies)
 - Collect all movies rated ≥ 3.5 by any group member → Remove target item from candidates
2. Pareto Filtering (Reduce 1275 movies → 3-5 items)
 - Pre-filter: $rs \geq 0.5$, $rate \geq 3.5$ → 345 items
 - Top-50 by rating
 - Compute 4 metrics
 - └ Keep non-dominated → 3-5 items
3. GROW Phase (Build explanation)
 - Add items iteratively
 - Rebuild recommender
 - └ Stop when target disappears
4. PRUNE Phase (Remove redundancy)
 - Try removing each item
 - └ Keep only if necessary (Drop fairness below 0.67)
5. Output: Minimal fair explanation

Example: "If group had NOT watched Full Metal Jacket,
then Like Water for Chocolate would NOT be recommended"



Results - Success & Size

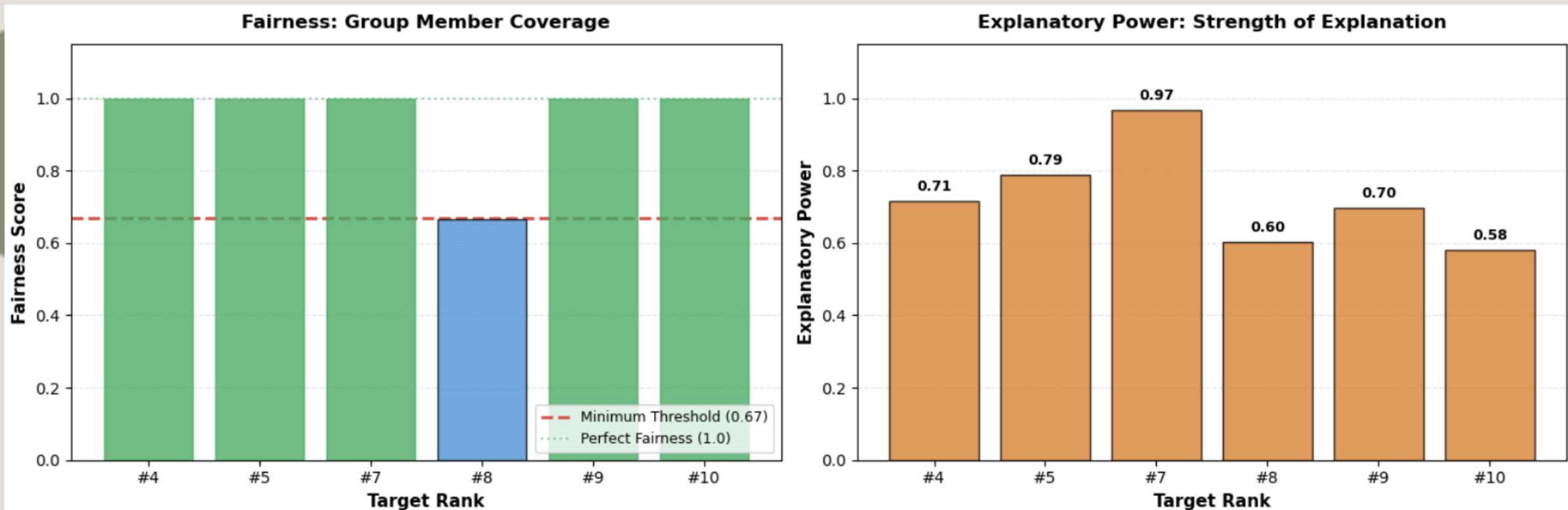


Rank	Target Movie	Size	Explanation Items (movieId)	Fairness	Exp Power
4	Hoop Dreams (1994)	3	[3275, 4011, 3949]	1.0	0.715
5	Return of the Pink Panther (1975)	2	[79132, 60069]	1.0	0.788
7	Being There (1979)	1	[48516]	1.0	0.967
8	Elf (2003)	1	[92259]	0.67	0.603
9	Gods Must Be Crazy (1980)	1	[1193]	1.0	0.697
10	Like Water for Chocolate (1992)	1	[1222]	1.0	0.580

Valid Explanations: 6/10 (60%)

Explanation Size:
Mean: 1.50 items
Range: 1 – 3 items

Results - Quality Metrics



Fairness Score:

Mean: 0.94

Perfect (1.0): 5/6

All above threshold (≥ 0.67): 5/6

Explanatory Power:

Mean: 0.725

Range: 0.580 – 0.967

