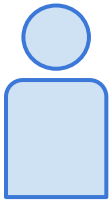
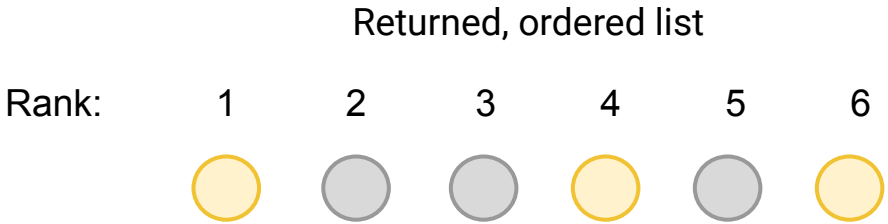
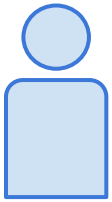
Matchmaking: Evaluation, ranked retrieval





Relevant item | Not relevant item

Matchmaking: Evaluation, ranked retrieval



Returned, ordered list

Rank (k):	1	2	3	4	5	6
Precision @ k:	1	0.5	0.33	0.5	0.4	0.5
Recall @ k:	0.33	0.33	0.33	0.66	0.66	1

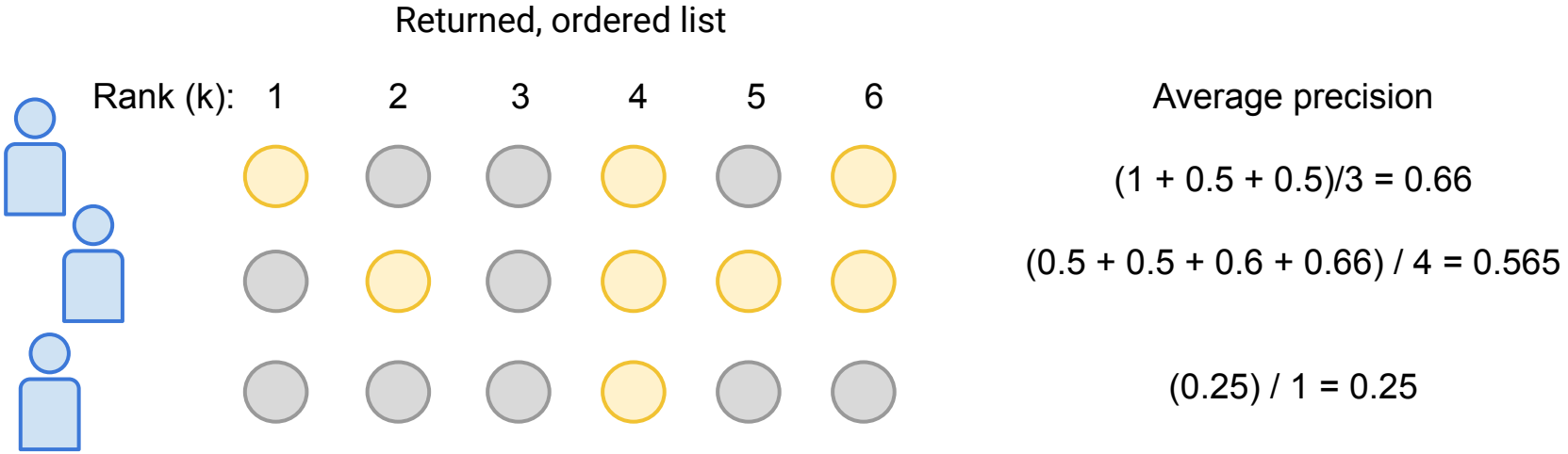
3 relevant items

Average precision: (1 + 0.5 + 0.5) / 3 = 0.66

Relevant item

Not relevant item

Matchmaking: Evaluation, mean average precision



Mean average precision: (0.66 + 0.565 + 0.25) / 3 = 0.492

Relevant item | Not relevant item

Matchmaking: Comparing models

	Model 1	Model 2	
AP@1	0.17	0.15	
AP@2	0.18	0.12	
AP@3	0.16	0.13	
AP@4	0.158	0.137	
mAP	0.167	0.134	**delta mAP = 0.033**

^ test statistic

Matchmaking: Comparing models

	Model 1	Model 2
AP@1	0.15	0.17
AP@2	0.18	0.12
AP@3	0.13	0.16
AP@4	0.158	0.137
mAP	0.1545	0.14675

delta mAP = 0.00775

Catalog for each shuffle