Personalized Recommendations of Upcoming Sport Events

Sebastian Cornelius Bergh

Motivation and problem statement

- Lots of **value** in the sports industry

- Separate research on linear shows and sports content

- **Gap** when it comes to recommending live sports events

- Several **challenges**
  - Live aspect
  - Balance between favorites and new content
  - Strong biases
  - Limited lifecycle

Motivation and problem statement

- Develop a novel **sports recommender system** for TV 2 Play.

- Little existing research

- Collaborative filtering to recommend upcoming live sports.

- Improve the **user experience**.

Research questions

- **RQ1:** Which collaborative filtering model is best suited for recommendation of sports content?

- **RQ2:** Which factors influence sports fans' preferences for specific types of live sports events?

Media
Futures ●

Datasets

- Received from TV 2
- 4 Different datasets
- Collected over approximately 3.5 months
- **Combined** based on different usage areas

Dataset	Observations	Features
Viewing sessions	> 1 000 0000	9
Metadata	6 838	7
Participants	1 768	8
Events	4 200	10

Media
Futures •

Exploratory analysis

- Early work focused on an exploratory analysis
- Gain a thorough **understanding** of the datasets
- Initial experiment focuses on **sports-type**
- Covered multiple aspects
  - Time
  - Items
  - Users

Datasets – exploratory analysis

- Start with K-means clustering
- Majority for additional information

User Id	Basketball watched	Cheerleading watched	E-sport watched	...	Fotball watched
1	13,000 sec	0 sec	5,000 sec	...	0 sec
2	0 sec	10,000 sec	0 sec	...	21,000 sec
3	0 sec	0 sec	5,000 sec	...	0 sec
4	0 sec	6,000 sec	7,000 sec	...	0 sec
5	0 sec	3,000 sec	0 sec	...	8,000 sec
6	0 sec	0 sec	0 sec	...	16,000 sec
7	15,000 sec	0 sec	0 sec	...	0 sec
8	0 sec	10,000 sec	6,000 sec	...	0 sec
9	0 sec	0 sec	0 sec	...	4,000 sec

Media Futures

Exploratory analysis - Users

• Group users to identify patterns and similarities
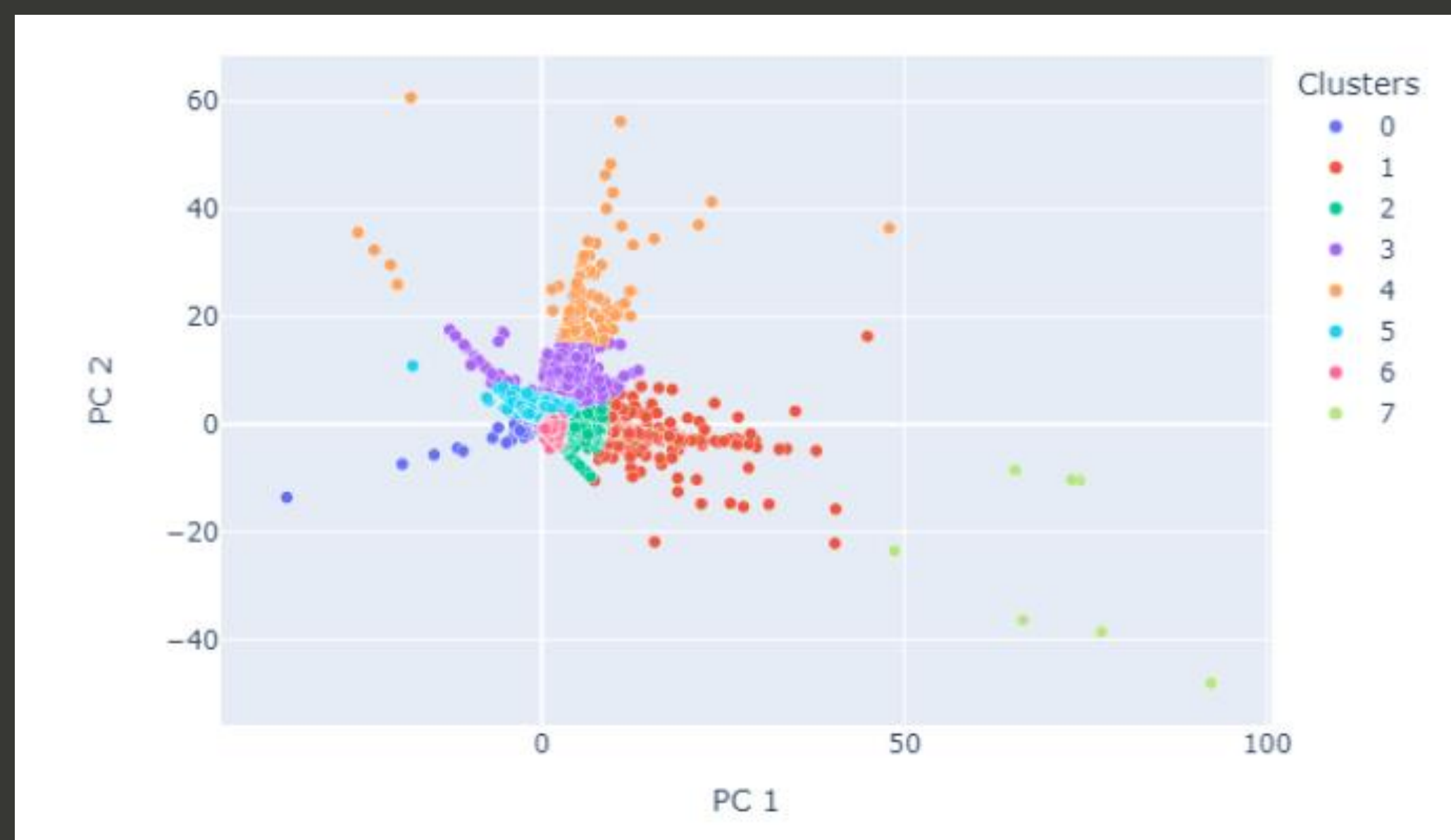
Cluster	Amount of users in % of total users	Favourites (% of users in cluster)	Second Favourites (% of users in cluster)	Average viewing time on favourites
Cluster0	69.1%	Football (67.9%)	None (78.6%)	1.53 h
		Sykkel (10.8%)	Football (7.6%)	0.49 h
Cluster1	2.4%	Ishockey (97.7%)	None (42.1%)	15.1 h
		Fotball (0.6%)	Fotball (34.5%)	1.68 h
Cluster2	1.5%	Fotball (99.9%)	None (38.4%)	48.39 h
		Håndball (0.1%)	Sykkel (24.7%)	0.51 h
Cluster3	0.8%	Sykkel (99.4%)	Fotball (44.0%)	38.29 h
		Sjakk (0.2%)	None (34.0%)	1.73 h
Cluster4	0.3%	Sjakk (98.5%)	Fotball (41.9%)	52.25 h
		Sykkel (0.6%)	None (26.5%)	3.39 h
Cluster5	19.4%	Fotball (97.2%)	None (65.1%)	9.58 h
		Sykkel (1.2%)	Sykkel (12.5%)	0.45 h
Cluster6	6.5%	Fotball (99.6%)	None (52.1%)	23.05 h
		Sjakk (0.1%)	Sykkel (18.5%)	0.35 h
Cluster7	0.1%	Fotball (99.5%)	None (32.0%)	118.2 h
		Sykkel (0.5%)	Sykkel (29.1%)	4.92 h

Datasets - models

- Measure item similarity
- Model building

Dataset	Users	Items	Interactions
Sport	> 100 000	20	> 400 000

User Id	Sport	Watch time in seconds
4	Fotball	17,000 sec
4	Sjakk	5,000 sec
4	Sykkel	9,000 sec
11	Sykkel	6,000 sec
11	Basketball	10,000 sec
25	Fotball	9,000 sec
25	Sjakk	7,000 sec
25	Sykkel	11,000 sec
25	Basketball	2,000 sec
25	Ishockhey	4,000 sec
25	Poker	20,000 sec
25	MyGame	3,000 sec

Media Futures

Exploratory analysis - items

- Sports-type items
- Identified by CF
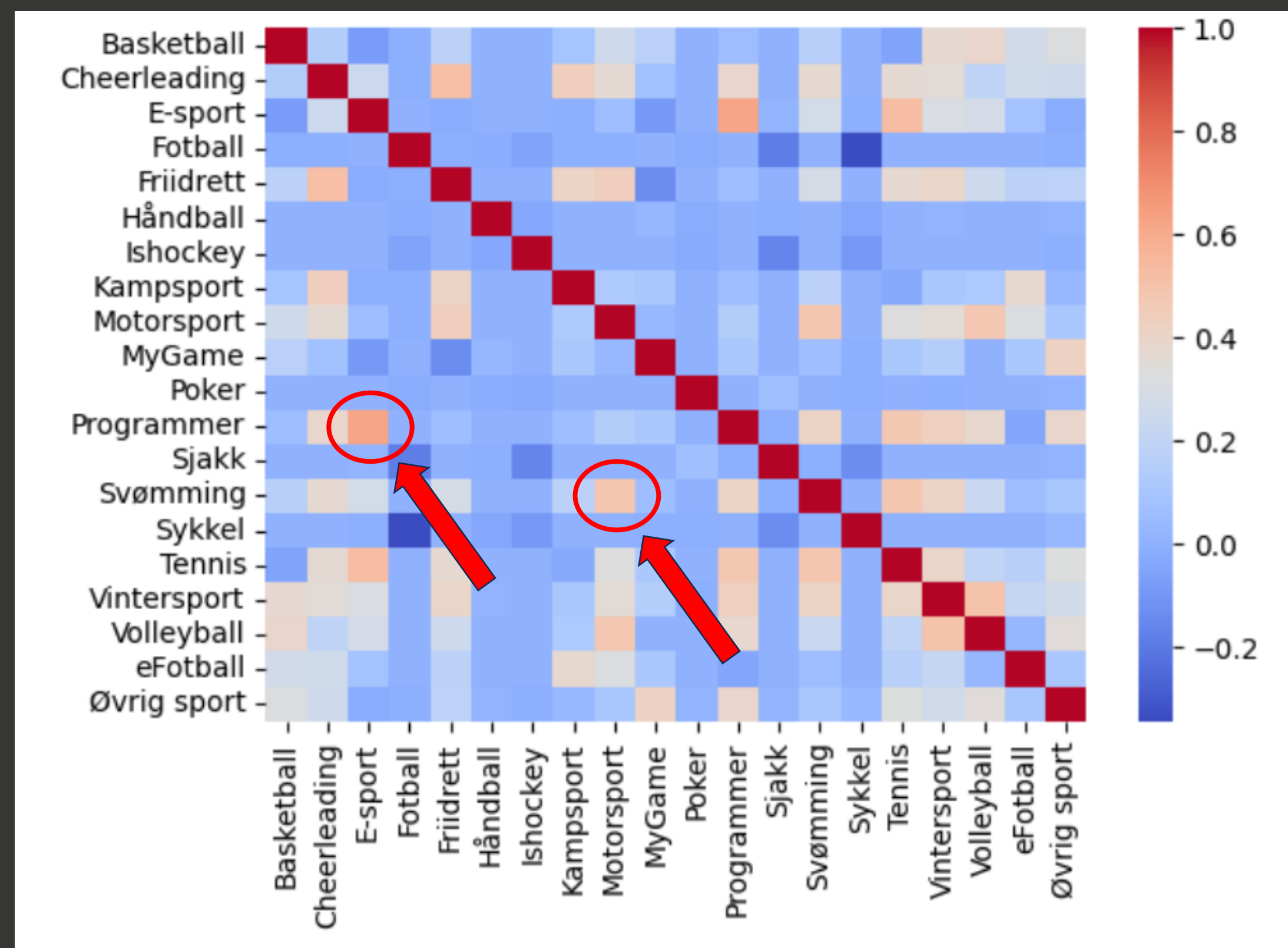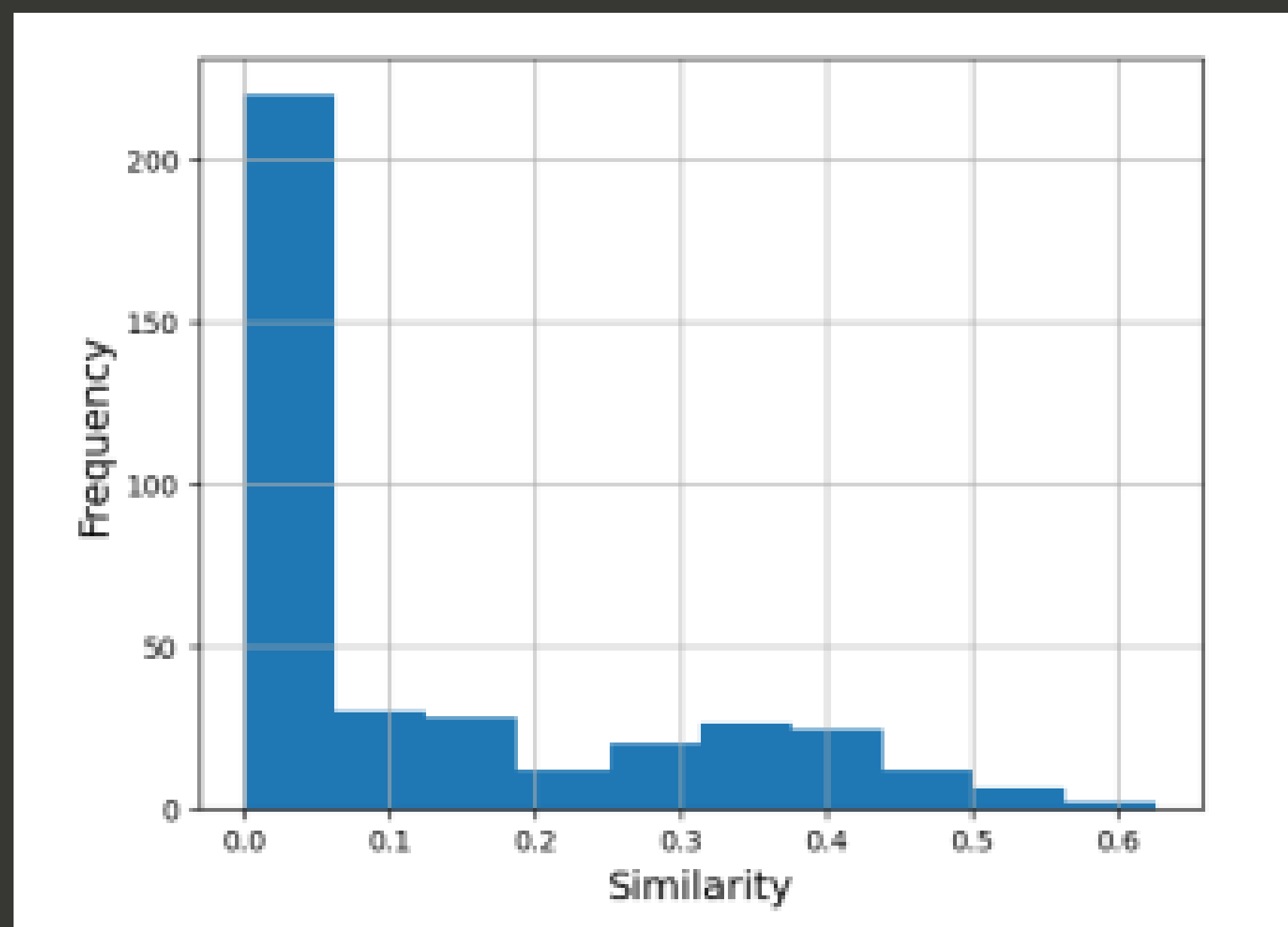
Offline evaluation – Sport models

- **Accuracy** and **beyond-accuracy** metrics
- ALS compared to baselines
- Take it **online**

Recommender algorithm	Type
Alternating Least Squares (ALS)	Pure CF
Bayesian Personal Ranking (BPR)	Pure CF
Logistic Matrix Factorization (LMF)	Pure CF
Most Popular	Non-personalized
Random	Non-personalized

Evaluation Metrics	Random	Popularity	ALS	BPR	LMF
P@10	0.054	**0.102**	0.091	0.061	0.102
R@10	0.532	**0.996**	0.892	0.600	0.996
AP@10	0.198	**0.732**	0.617	0.370	0.461
NDCG@10	0.275	**0.797**	0.684	0.423	0.590
Hit@10	0.538	**0.997**	0.894	0.607	0.996
RR@10	0.201	**0.735**	0.620	0.374	0.463
ROC_AUC	0.501	**0.938**	0.836	0.565	0.880
PR_AUC	0.230	**0.733**	0.623	0.397	0.461

Media Futures
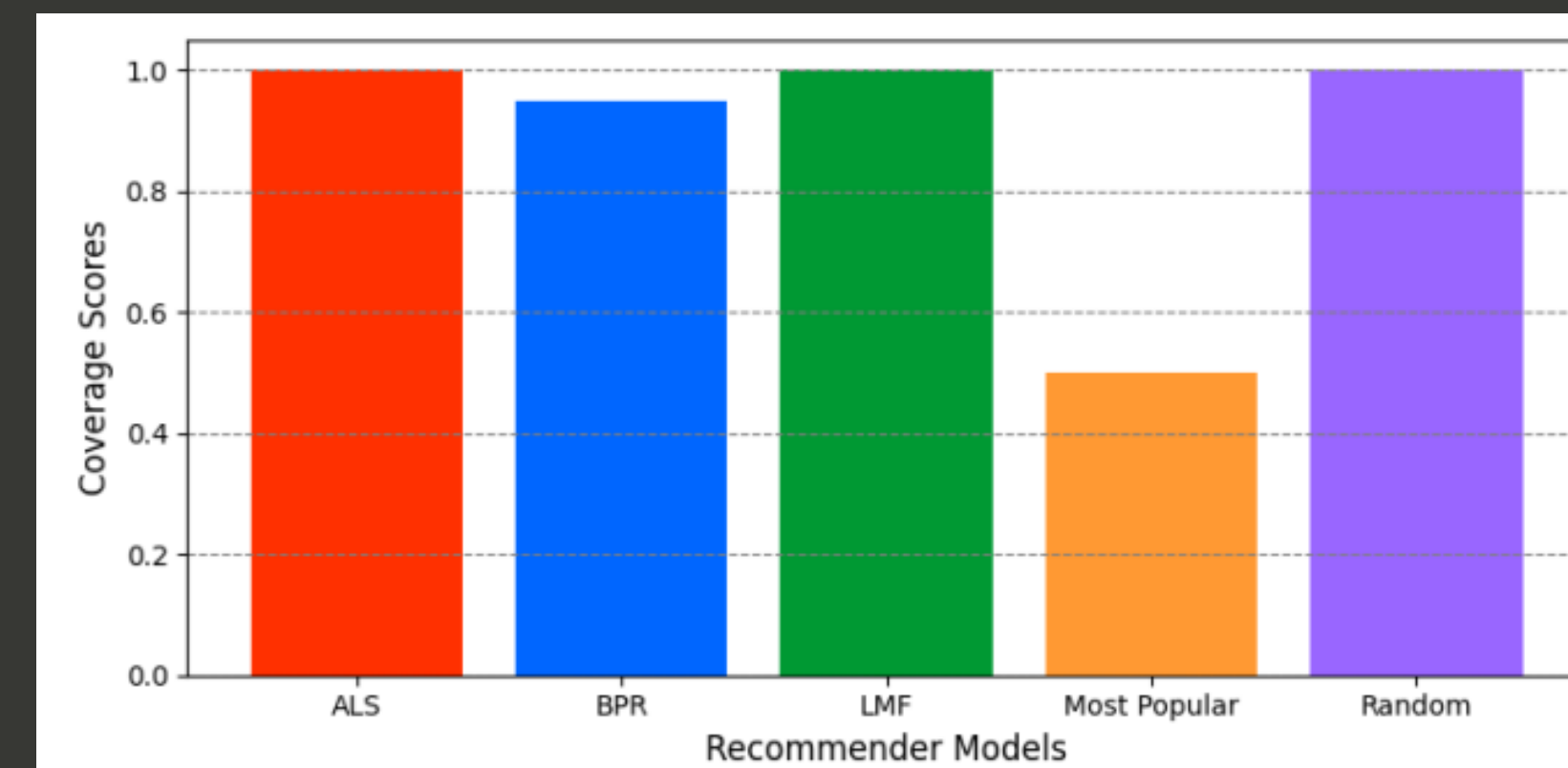
A/B testing – Online experiment 1

- Personalize the sports («Idretter») feed
- Displays a variety of sports
  - Redirects users to a dedicated page for each sport
- TV 2's current approach
- From March 30th to April 12th
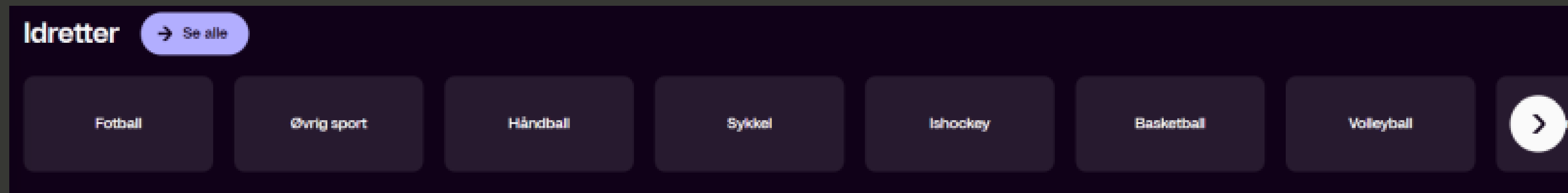
A/B testing – Results

- Baseline slightly outperforms CF in CTR
- Other factors to explore
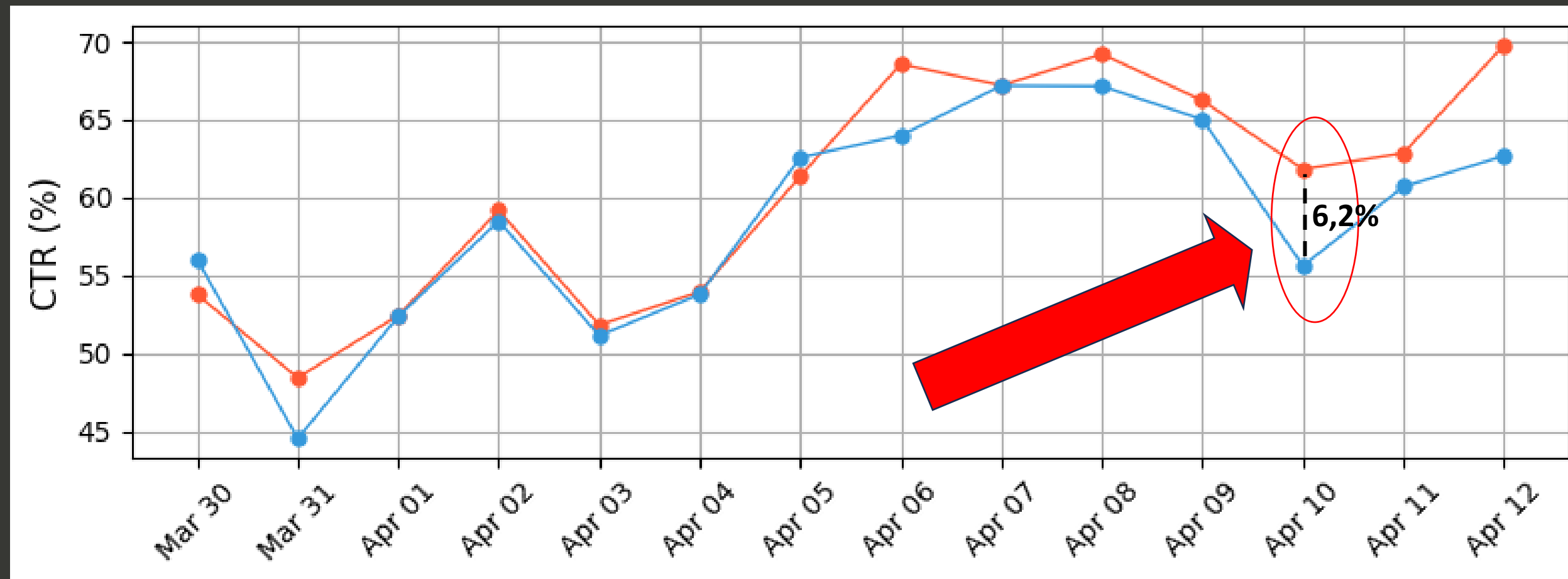
Metric	CF (ALS)	Time-based
Views	21,560	18,489
Clicks	11,878	10,397
CTR	55,09%	56,23%

Further development

- Take it one step further

- Another experiment

- Include tournaments and teams

Media
Futures •
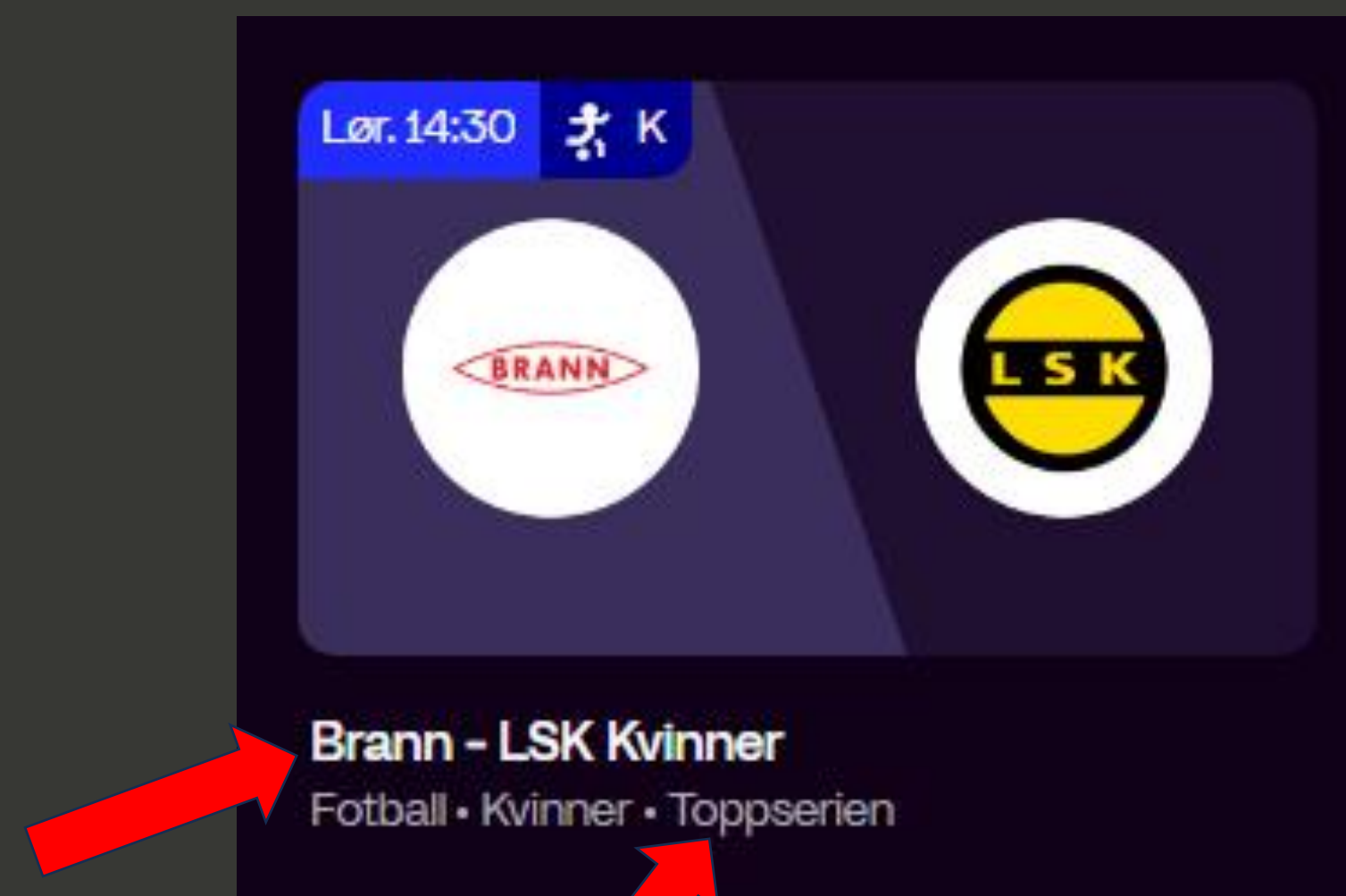
Exploratory analysis - Users

Users clustered by tournament
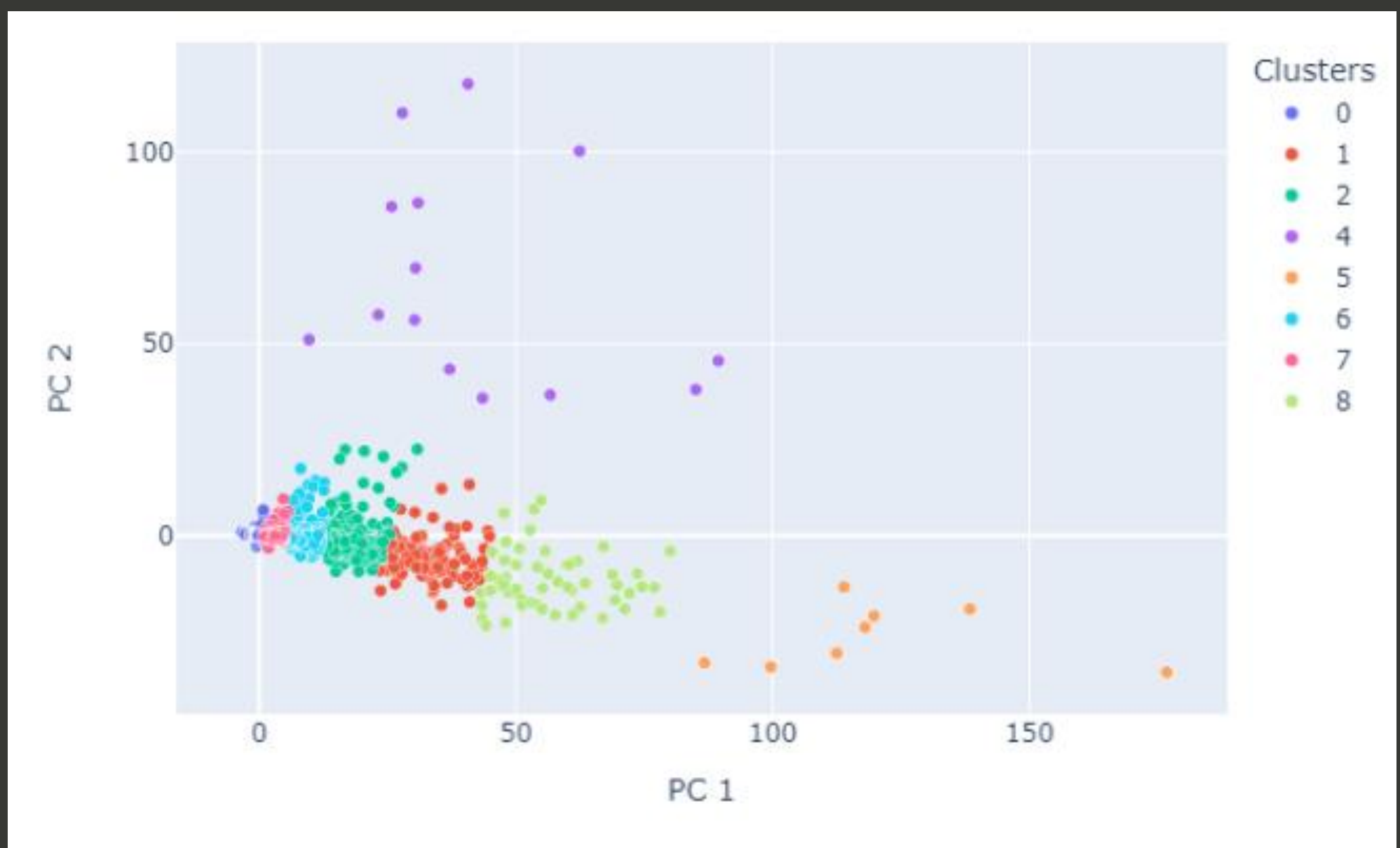
Cluster	Amount of users in % of total users	Favourites (% of users in cluster)	Second Favourites (% of users in cluster)	Average viewing time on favourites
Cluster0	14.38%	UEFA Champions League - Fotball (92.75%) Eliteserien - Fotball (2.02%)	UEFA Nations League A - Fotball (23.72%) None (22.43%)	5.96h 0.68h
Cluster1	70.36%	UEFA Champions League - Fotball (28.21%) UEFA Nations League B - Fotball (14.51%)	None (54.82%) UEFA Champions League - Fotball (9.75%)	0.58h 0.34h
Cluster2	4.72%	Eliteserien - Fotball (97.31%) UEFA Champions League - Fotball (1.66%)	UEFA Champions League - Fotball (46.12%) UEFA Nations League A - Fotball (11.87%)	11.15h 2.64h
Cluster3	2.89%	Eliteserien - Ishockey (99.07%) 1. divisjon, menn - Ishockey (0.23%)	None (41.16%) UEFA Champions League - Fotball (9.02%)	14.12h 0.27h
Cluster4	0.001%	Eliteserien - Fotball (100%)	UEFA Champions League - Fotball (66.67%) Norsk 1. divisjon - Fotball (33.33%)	524.93h
Cluster5	0.95%	Eliteserien - Fotball (96.61%) Norsk 1. divisjon - Fotball (1.60%)	UEFA Champions League - Fotball (46.48%) Norsk 1. divisjon - Fotball (23.74%)	30.51h 5.22h
Cluster6	2.75%	UEFA Nations League A - Fotball (88.76%) UEFA Champions League - Fotball (4.67%)	UEFA UEFA Champions League - Fotball (33.98%) UEFA Nations League B - Fotball (24.33%)	9.20h 2.83h
Cluster7	1.69%	Norsk 1. divisjon - Fotball (96.70%) Eliteserien - Fotball (1.47%)	UEFA Champions League - Fotball (37.90%) Eliteserien - Fotball (27.14%)	13.63h 2.53h
Cluster4	2.23%	UEFA Champions League - Fotball (92.79%) UEFA Nations League A - Fotball (3.71%)	UEFA Nations League A - Fotball (43.17%) Eliteserien - Fotball (22.53%)	16.23h 5.05h

Datasets - models

- Measure item similarity
- Model building
- More items

User Id	Tournaments	Watch time in seconds
4	Eliteserien - Ishockey	17,000 sec
4	1. divisjon - Fotball	5,000 sec
4	UEFA Nations League	9,000 sec
11	Eliteserien - Fotball	6,000 sec
11	Eliteserien - Ishockey	10,000 sec
25	UEFA Champions Leauge - Håndball	9,000 sec
25	UEFA Nations League - Fotball	7,000 sec
25	1. divisjon - Fotball	11,000 sec
25	LaLiga - Fotball	2,000 sec
25	Eliteserien - Håndball	4,000 sec
29	UEFA Nations Legue - Fotball	20,000 sec
29	1. divisjon - Ishockey	3,000 sec

Dataset	Users	Items	Interactions
Tournament	> 100 000	195	> 600 000

Media Futures

Exploratory analysis - items

- Tournament items
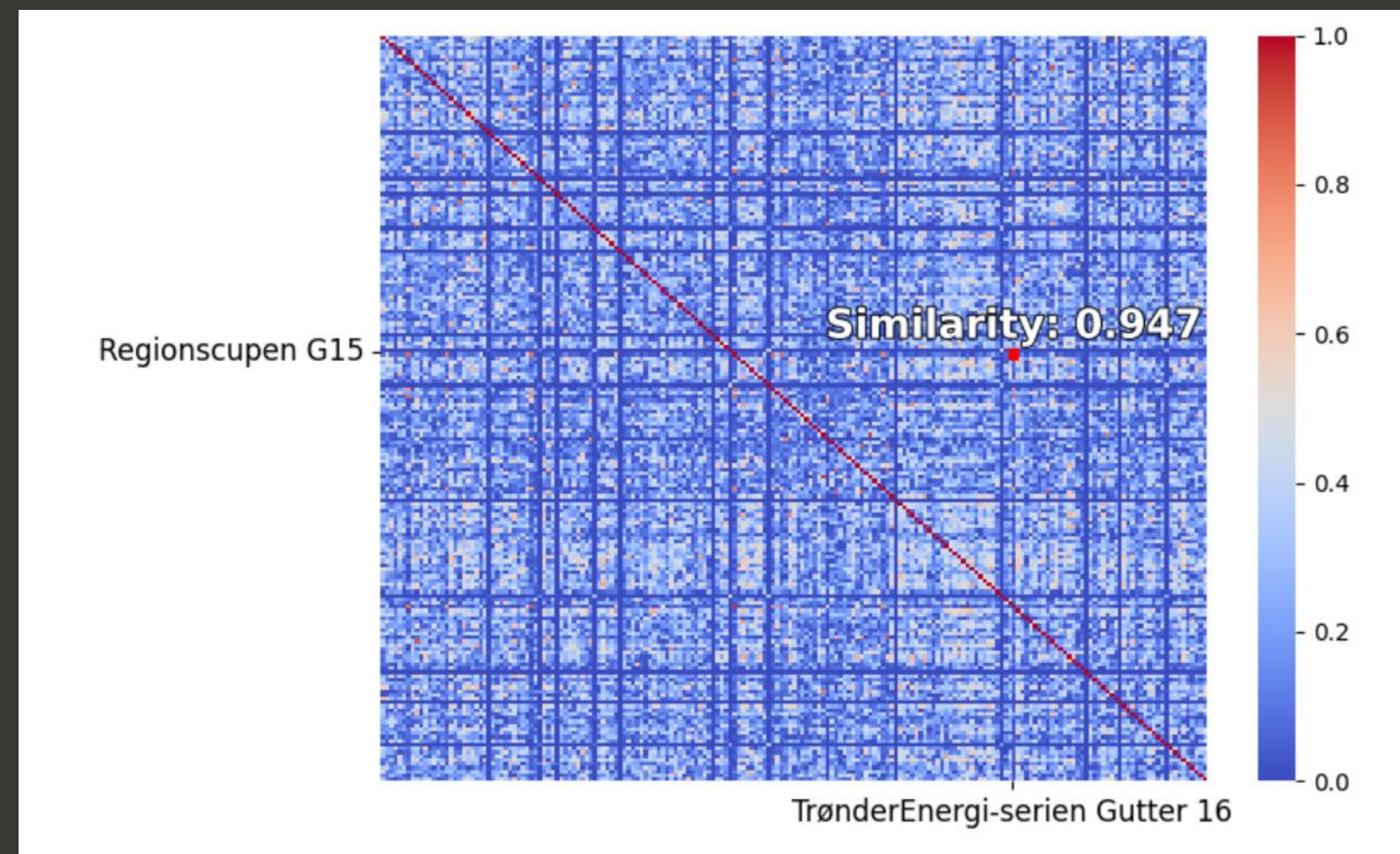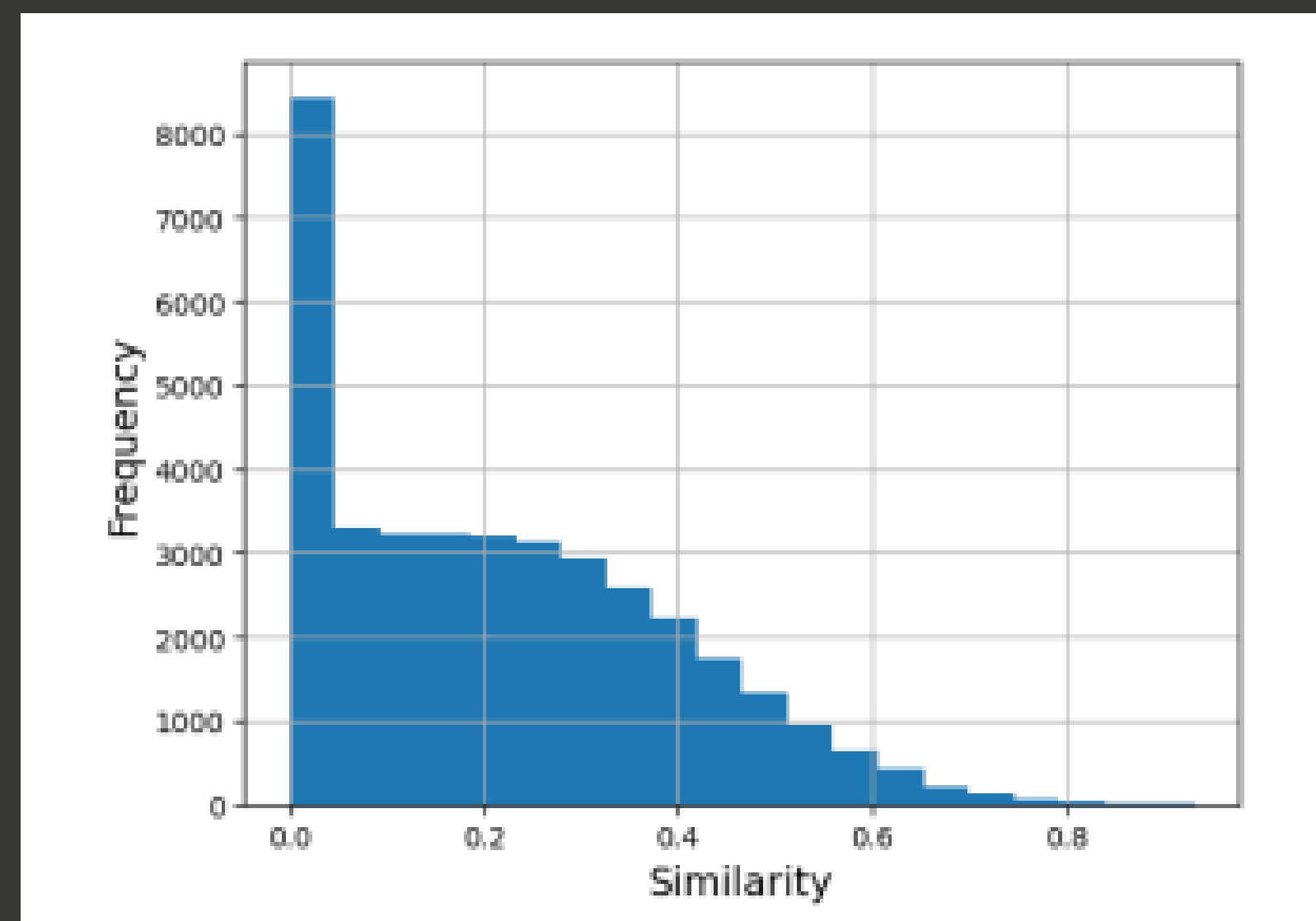
Offline evaluation – Tournament models

- Accuracy and beyond-accuracy metric
- Models trained on tournament data
- Take it **online**

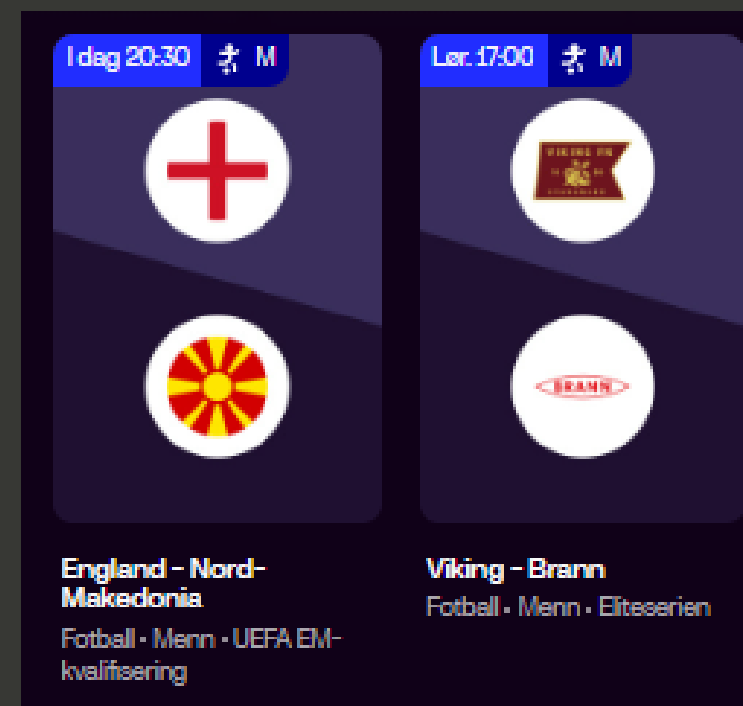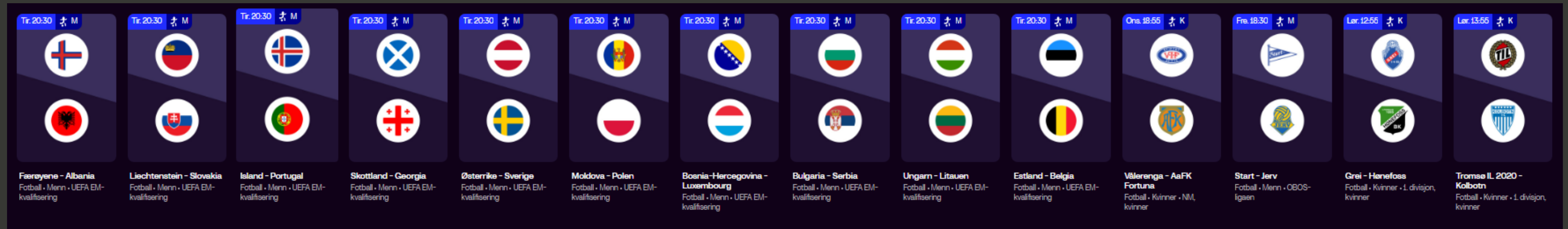| Evaluation Metrics	Random	Popularity	ALS	BPR	LMF
P@10	0.003	**0.110**	0.104	0.047	0.109
R@10	0.030	**0.909**	0.862	0.403	0.904
AP@10	0.005	**0.576**	0.521	0.305	0.360
NDCG@10	0.011	**0.664**	0.611	0.341	0.492
Hit@10	0.037	**0.931**	0.887	0.450	0.929
RR@10	0.006	**0.607**	0.550	0.345	0.377
ROC_AUC	0.501	**0.981**	0.936	0.608	0.969
PR_AUC	0.022	**0.583**	0.529	0.313	0.367

Media
Futures
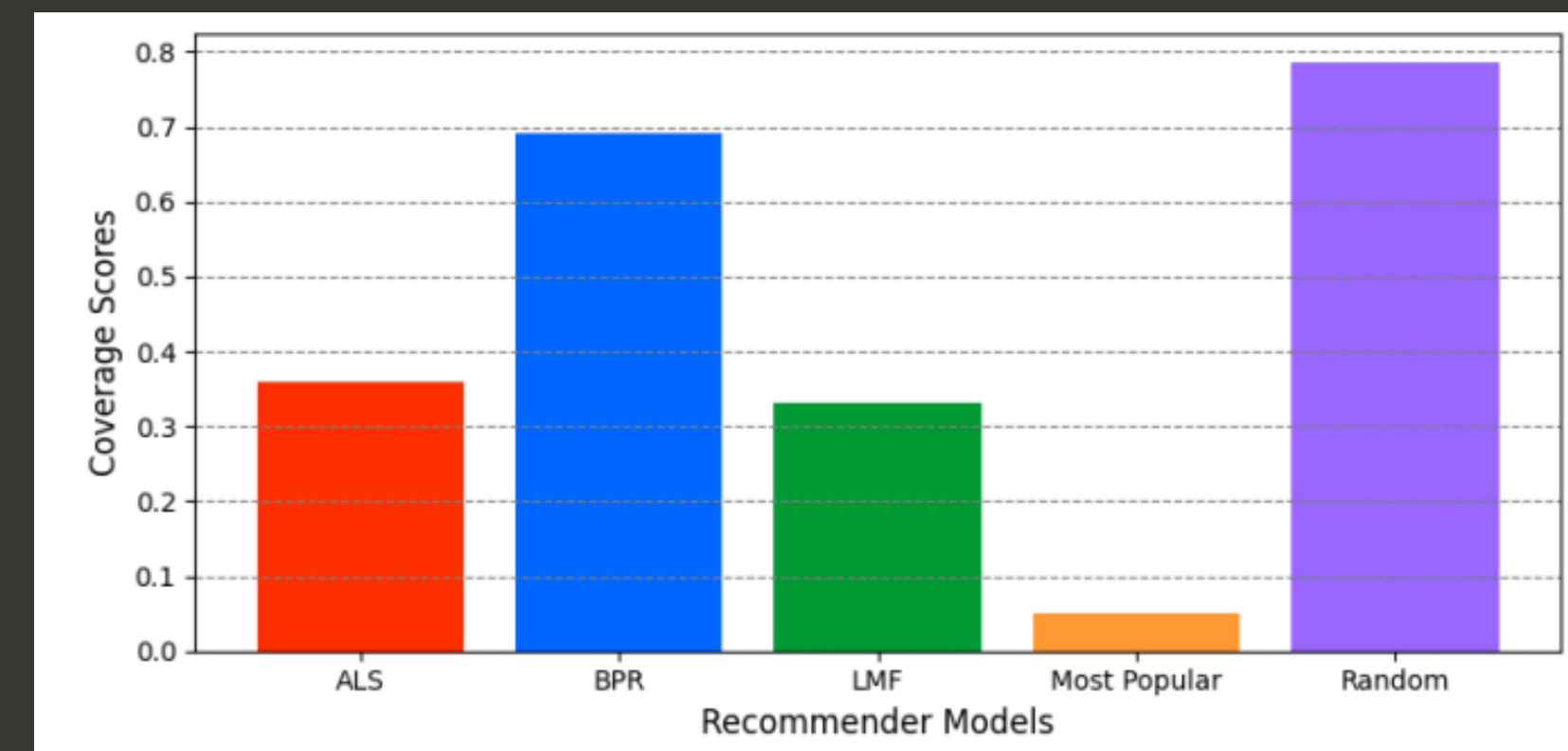
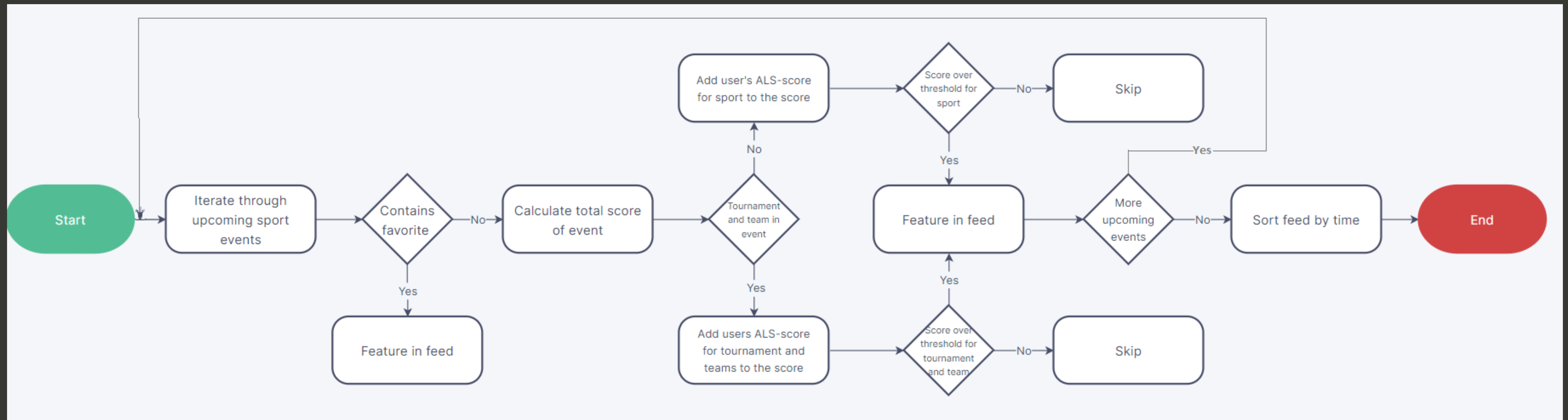# A/B testing – Online experiment 2

- Personalize the sport for you for the next seven days («Sport for deg neste 7 dager») feed
- TV 2's current approach
- A/B test from May 12th to May 28th

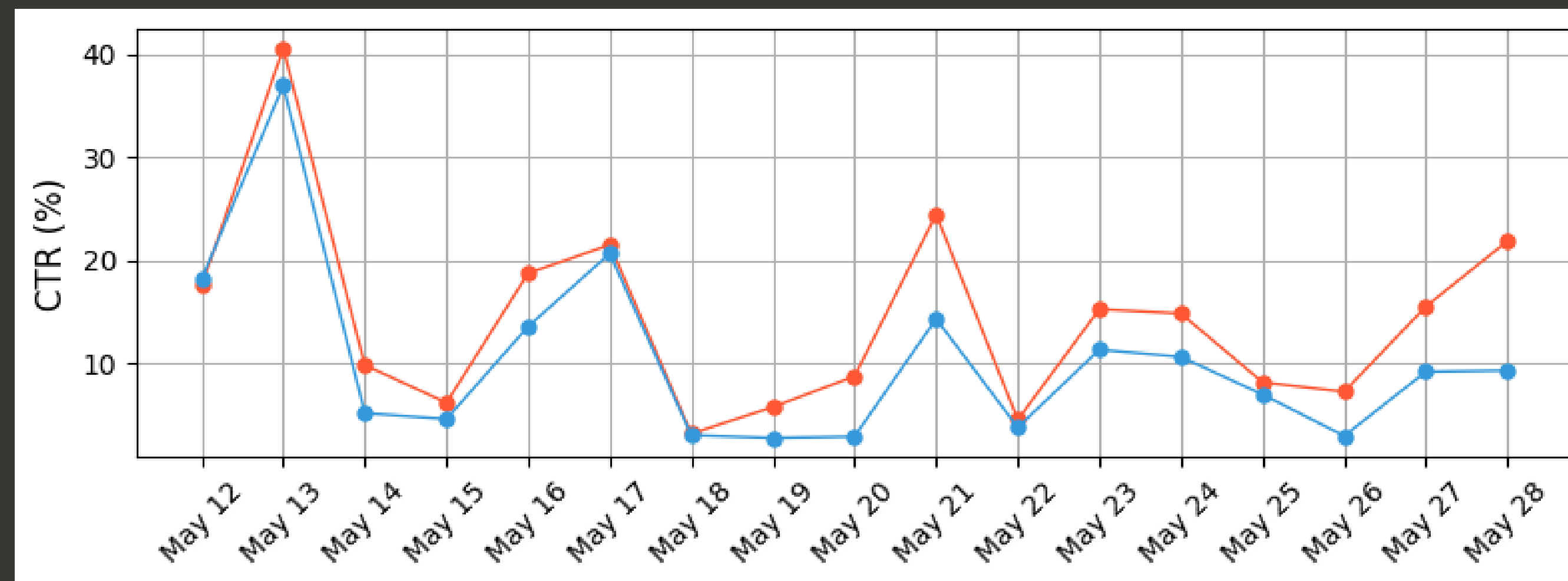# A/B testing – Online experiment 2

A/B testing – Online experiment 2

A/B testing – Results

- Baseline outperforms CF in CTR
- Generates more views
- Other factors to explore

Metric	CF (ALS)	Implicit favorites
Views	44,767	35,281
Clicks	6,661	6,745
CTR	14,88%	19,12%

Media
Futures

Conclusion

- Novel CF approach for sports events
- Offline evaluation
- A/B tests

Media Futures ●

Challenges and future work

- Not a lot of research in this field

- Potential of **publication**

- Sport aspect

- Limited analysis of A/B tests
  - Explore additional factors

- Include teams in exploratory analysis

- Further exploration with online experiment 2

Media
Futures