ReTV

Re-Inventing TV for the Interactive Age

AI for Audience Prediction and Profiling

to power innovative TV Content Recommendation Services

Lyndon Nixon, MODUL Technology

Krzysztof Ciesielski, Genistat AG

Basil Philipp, Genistat AG

AI4TV @ ACMMM2019, 21/10/2019

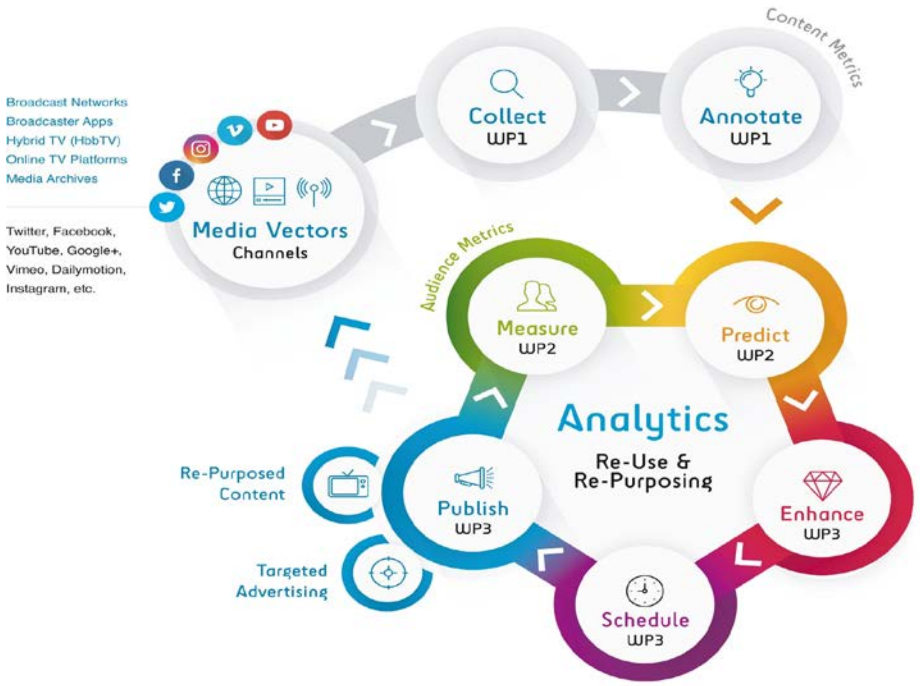
ReTV

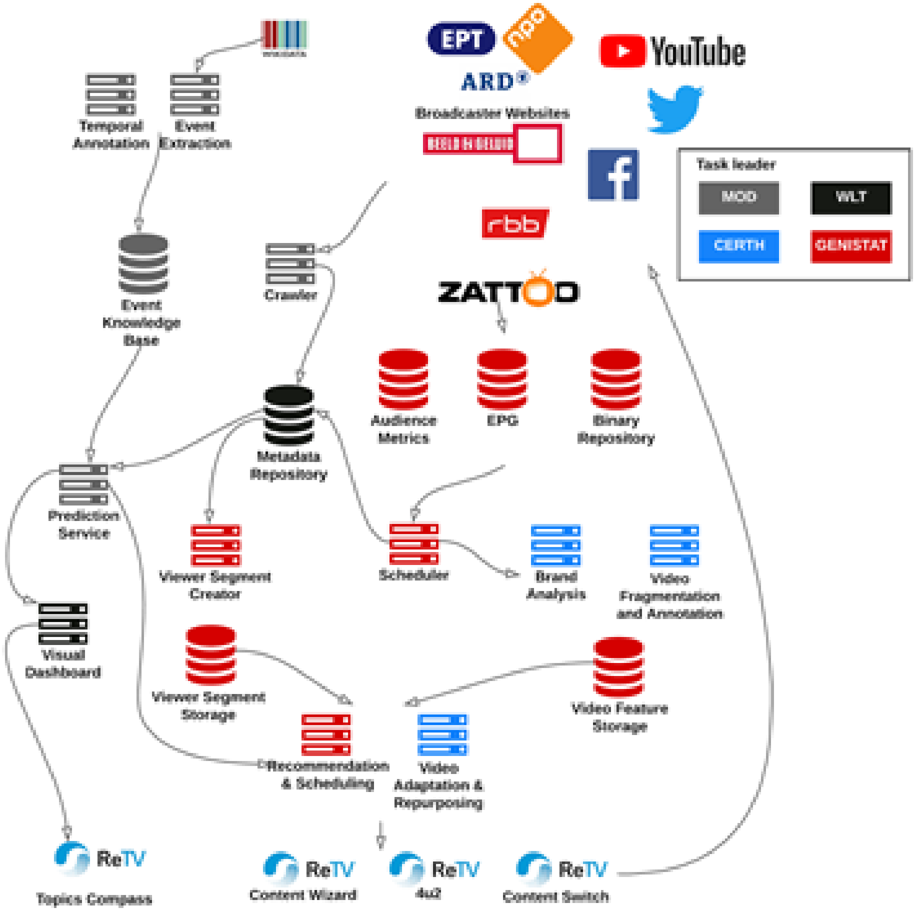
Viewing of linear broadcast TV is decreasing while time spent with digital content on Catchup TV, on-demand OTT or social media rises.

Broadcaster audiences are fragmented across digital channels and digital channels are full of competing content offers for their limited attention.

The TV industry is still catching up with their online competition in the use of Web technology: user tracking, personalisation and targeting.

ReTV will develop a Trans-Vector Platform (TVP) to analyse content across all channels and "publish to all media vectors with the effort of one"





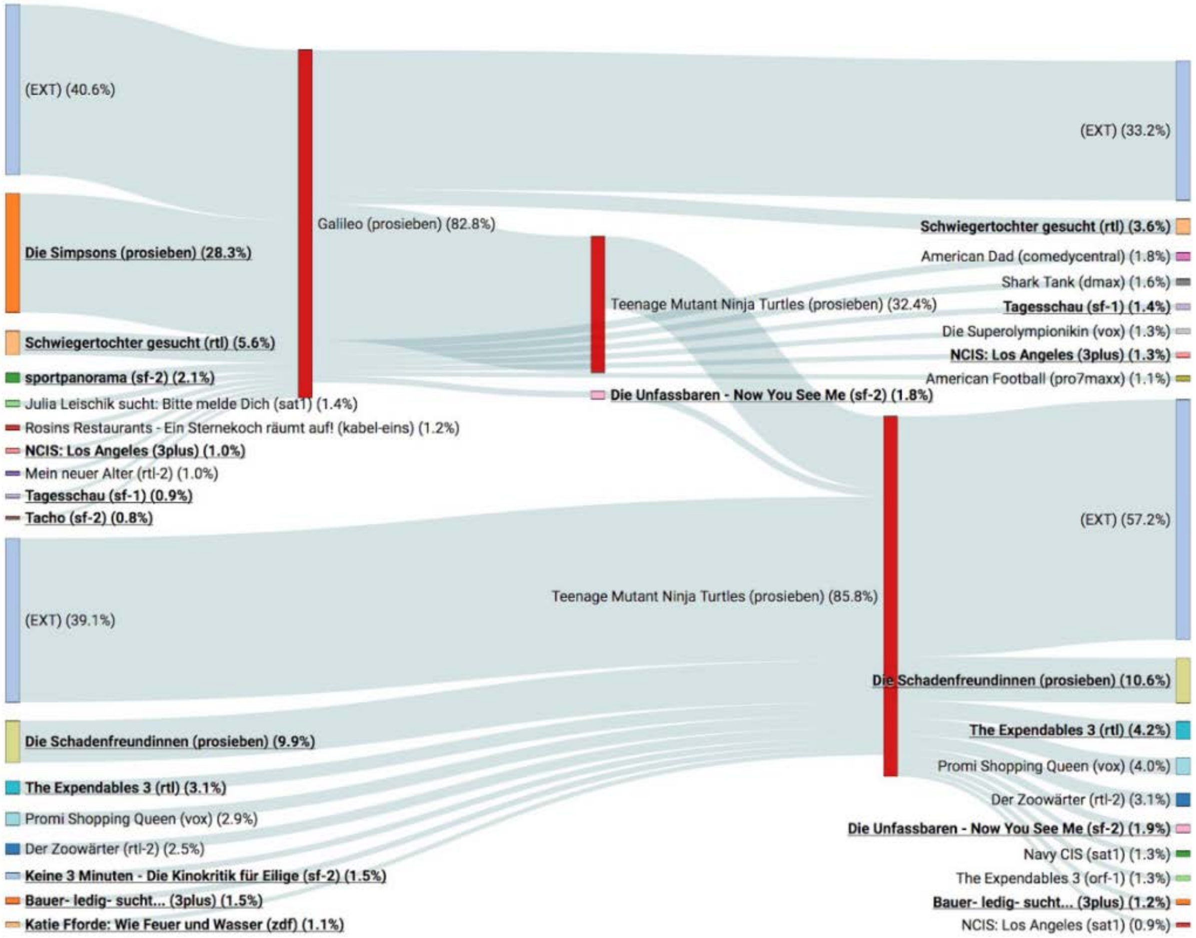
ReTV

Audience Metrics

TV Audience is a key success metric for TV content creators and distributors.

Prediction of Audience is used to determine TV content value, e.g. for pricing of TV advertising or ROI of licensing content for broadcast.

Audience is a characteristic of and a dependent variable of the TV programming. Our metrics aggregate total viewers per channel every 5 minutes with smoothed averaging.



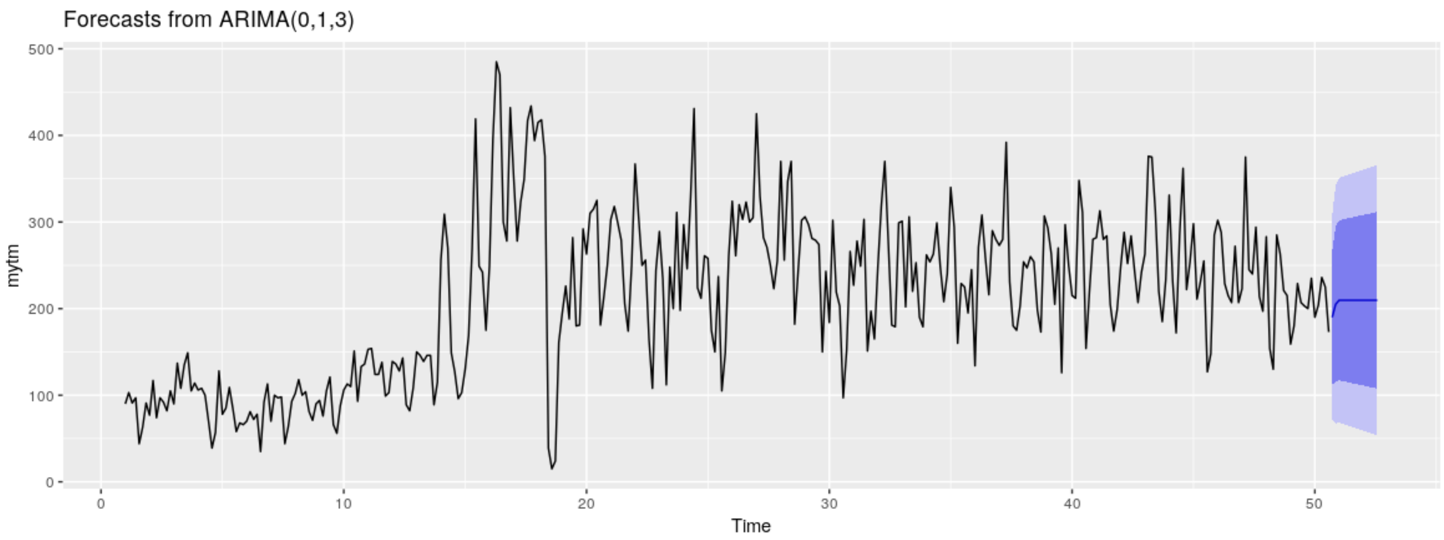
ReTV

Audience Prediction

TV Audience prediction relies heavily on the seasonality and trend components of past audience, as a result of regularity in TV programming (e.g. weekly schedule, summer and fall seasons).

Statistical forecasting methods like ARIMA can be used.

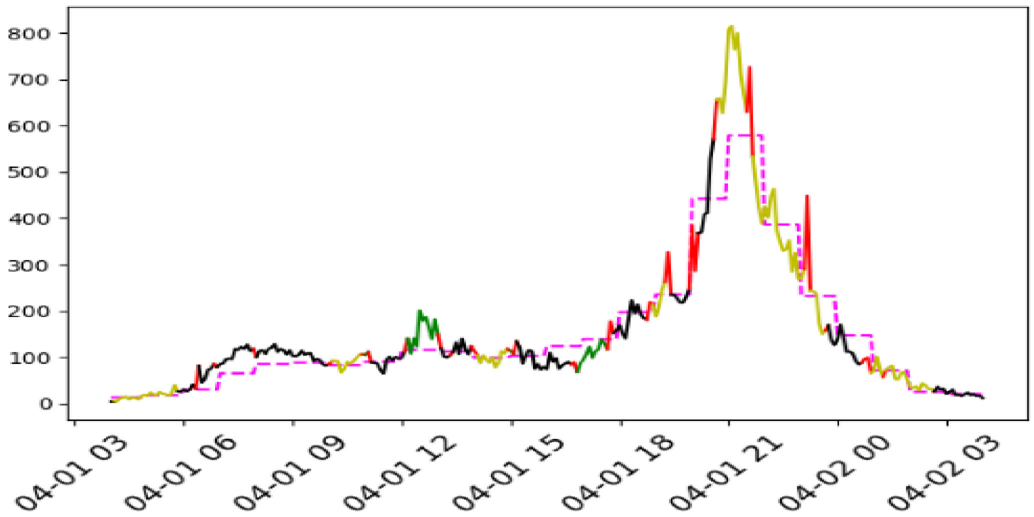
However, these models consider the irregular data as **outliers** and exclude them / smooth them out.



Prediction: the role of content type

Our baseline audience prediction used **random forest** models on viewing numbers per TV channel.

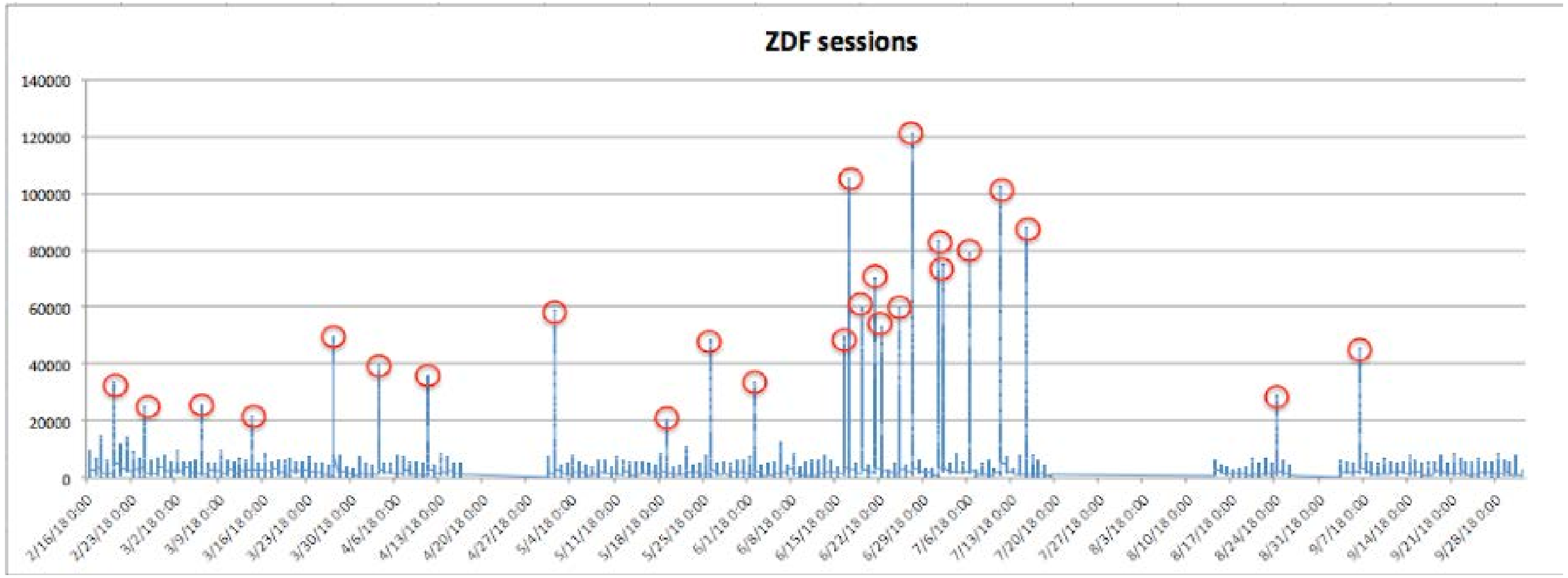
Noting how drops and rises in viewers correlated to start and end of ad breaks, we worked on aligning the time-series based audience data with our EPG data, using program categories.



Audience: the meaning of outliers

Audience data outliers can be meaningful, and provide features for more accurate prediction.

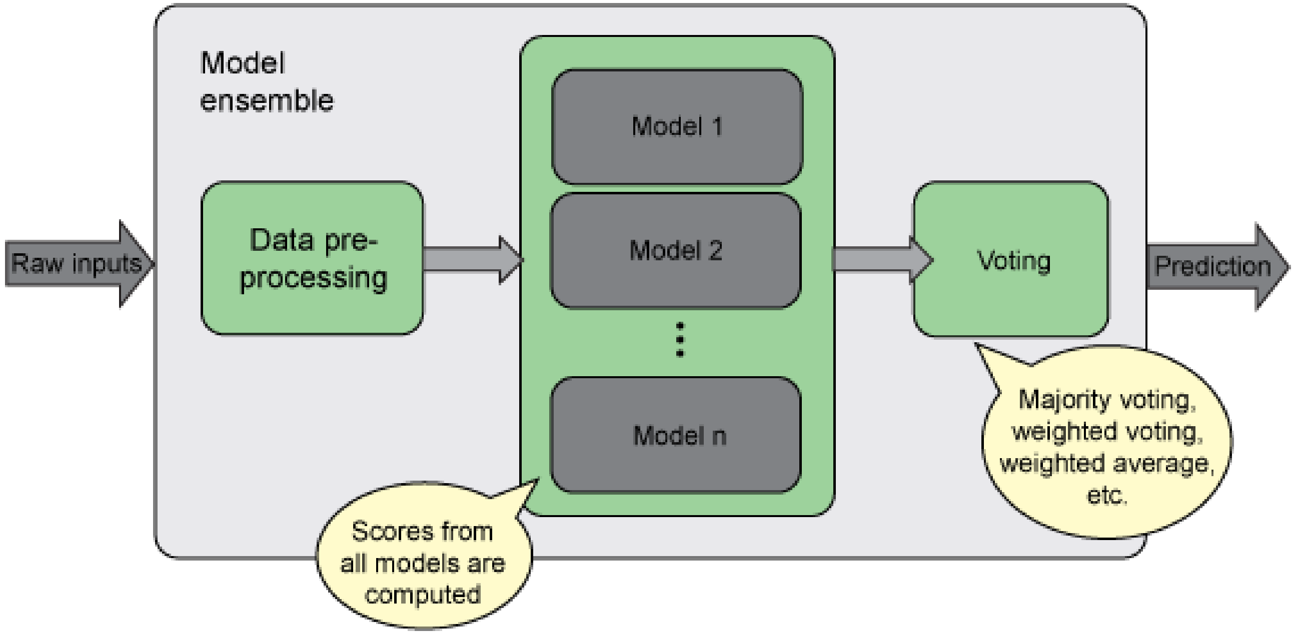
We studied outliers in our audience data and found 100% were related to the broadcast of external events, and 94% were live sports broadcasts.



AI based Audience Prediction

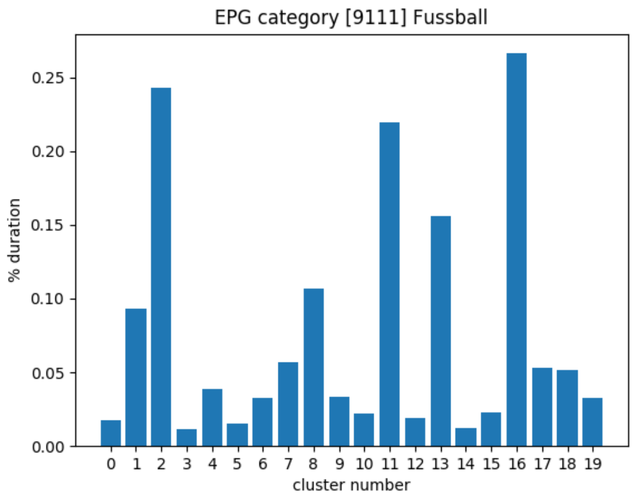
In AI methods (Machine Learning) we can train a prediction model with selection of additional features.

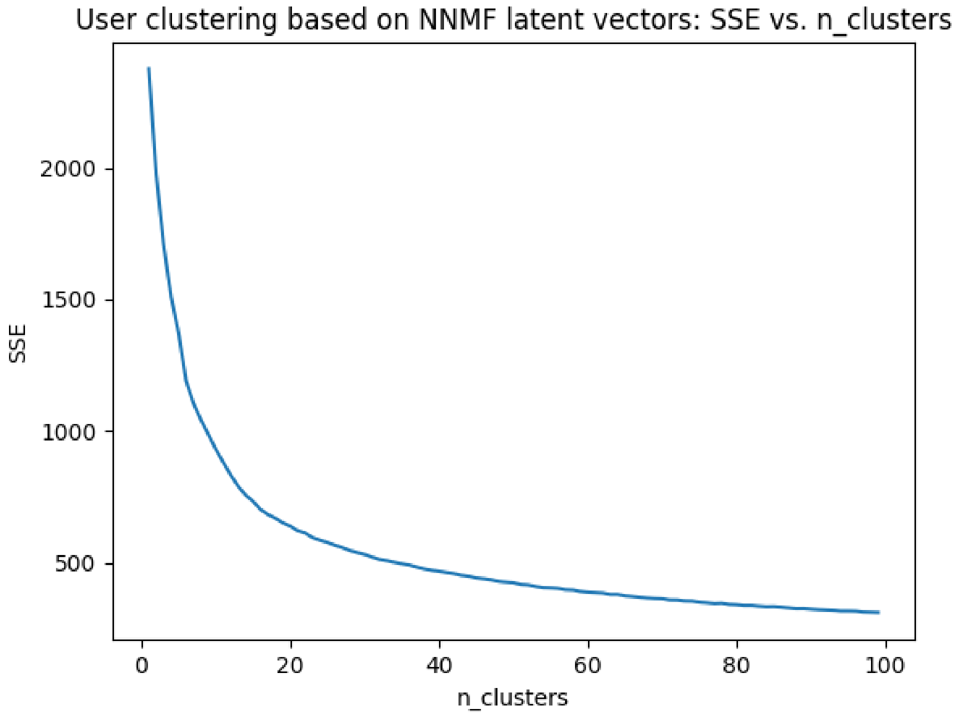
- Use topics of content as a feature (extracted from EPG data), not just the content (TV series) itself
- Use external events as a feature (account for irregularity)

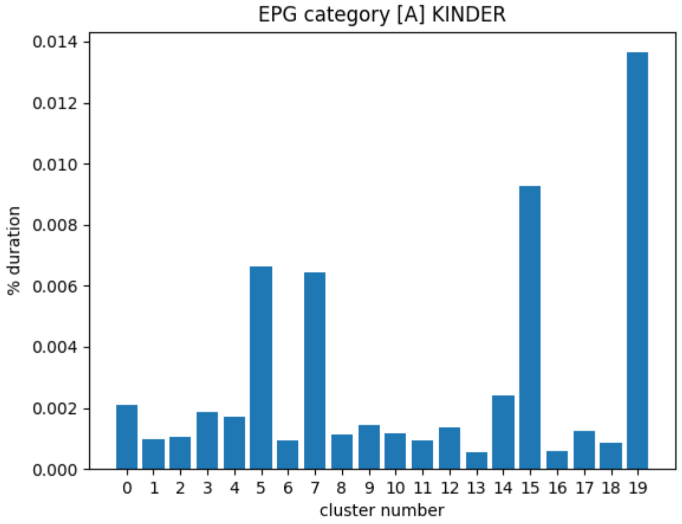


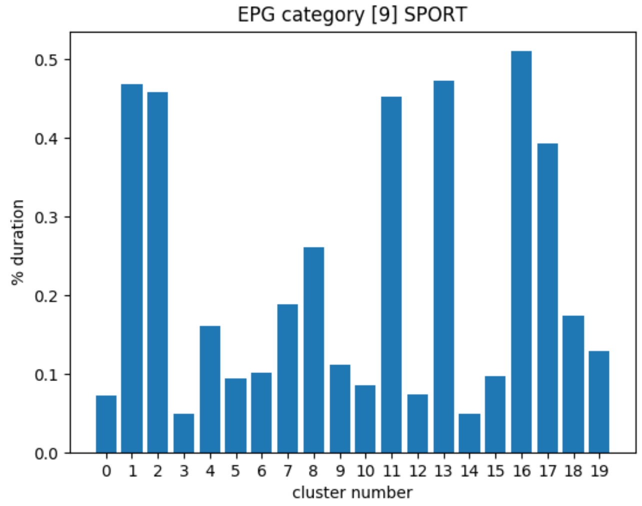
AI based Audience Segmentation

The same AI method can group viewers with similar viewing patterns into clusters. We found 80-100 clusters to be ideal to represent differing viewer preferences. The segments reflect a hierarchical structure of interests.









ReTV

Our AI approach

The starting point was Non-Negative Matrix Factorization (NNMF) – standard collaborative filtering using as input the interactions between viewers and content.

As an extension, we use Field-Aware Factorization Machines (FFM) which supports additional features. We trained prediction with content and event features.

Model	MAE	Pearson corr.	Spearman corr.
NNMF	0.45	0.32	0.35
FFM (no EPG)	0.25	0.73	0.75
FFM (coarse EPG)	0.22	0.75	0.77
FFM (detailed EPG)	0.18	0.79	0.8

SRF-info		lag / events			lag / no_events			no_lag / events			no_lag / no_events		
category	cnt	MAE	Spearman	Pearson	MAE	Spearman	Pearson	MAE	Spearman	Pearson	MAE	Spearman	Pearson
info	996	24.33	0.99	0.71	17.83	0.99	0.8	12.99	0.99	0.94	26.37	0.99	0.67
film	0	-	-	-	-	-	-	-	-	-	-	-	-
sport	**86**	**6.07**	**0.99**	**1**	**7.45**	**0.99**	**0.99**	**6.45**	**0.99**	**0.99**	**27.58**	**0.98**	**0.86**
ads	7134	15.13	0.99	0.87	15.02	0.99	0.87	17.77	0.99	0.8	16.77	0.99	0.87
other	3359	13.95	0.99	0.88	16.49	0.99	0.81	16.53	0.99	0.89	13.99	0.99	0.92
ALL	11576	15.51	0.99	0.84	15.63	0.99	0.84	16.91	0.99	0.83	16.87	0.99	0.84

Feature selection

Content and event characteristics were used as features and we looked at feature importance.

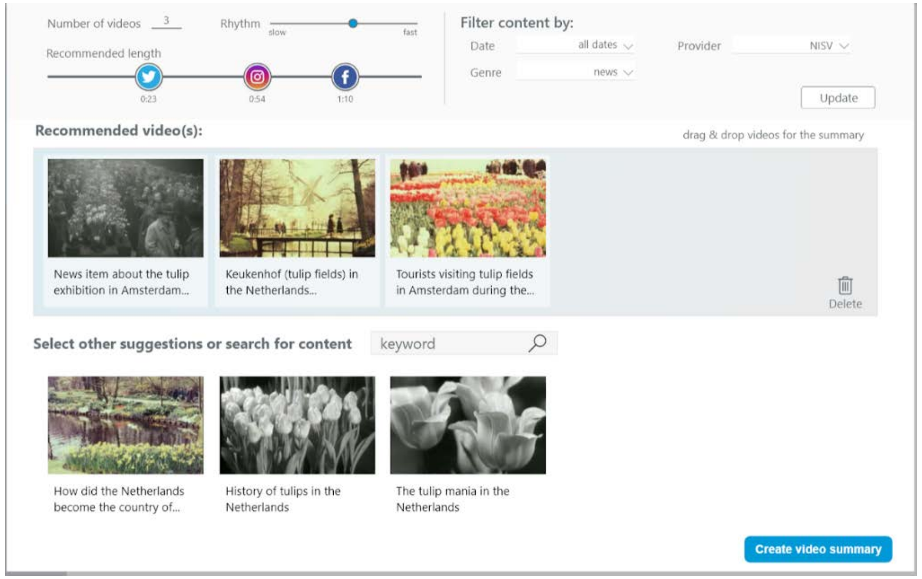
model with event features	weigh t	gain	cover	total_gain	total_cover
content_epg_main_cats_9	3	57,627,356	5,855	172,882,068	17,566
event_people_Rafael Nadal	2	26,638,854	7,328	53,277,708	14,656
event_stage_Finals	5	24,150,707.90	6,038	120,753,539	30,191
event_people_Roger Federer	1	16,545,830	7,333	16,545,830	7,333
event_countries_SUI	4	13,572,314	2,222	54,289,258	8,889
event_countries_SRB	3	13,345,356	6,756	40,036,068	20,268
event_countries_GER	1	9,317,020	6	9,317,020	6
event_people_Stan_Wawrinka	4	8,681,276	1,833	34,725,105	7,333
event_stage_Viertelfinal	2	8,278,286	17	16,556,572	34
content_duration	84	8,178,525	1,949	686,996,112	163,749

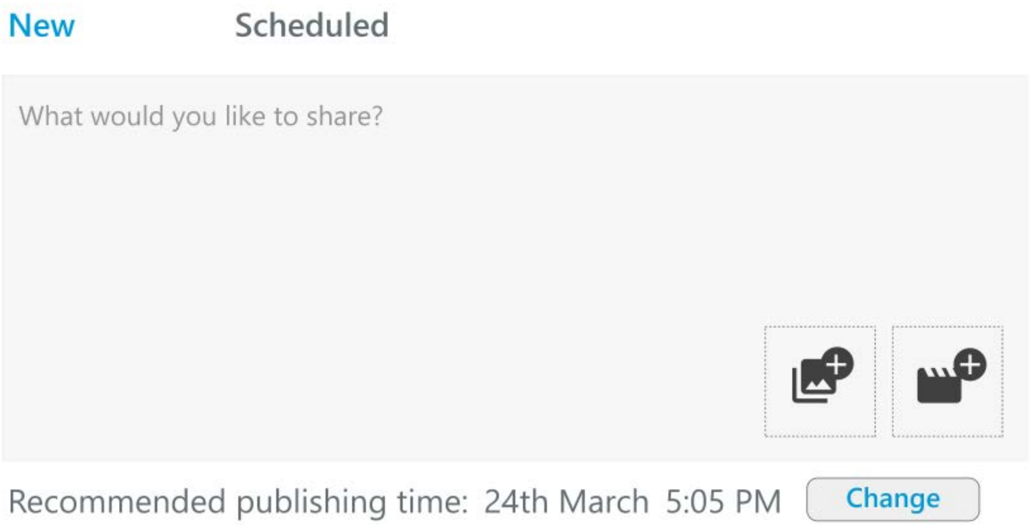
Content Recommendation & Scheduling

Prediction
- For a chosen content item, predict the audience at a future broadcast time

Recommendation
- Determine the most relevant content for this viewer in their viewing context

Scheduling
- Identify „most likely to be successful" time to publish content on this channel





ReTV

Content sWitch and 4U2 demos later today!

We have prototypes of 2 professional user tools: trending topic detection and content selection & re-purposing

The ReTV Stakeholder Forum is your opportunity to engage with us, be first to get updates and have the opportunity to test our tools and applications!

ReTV

Re-Inventing TV for the Interactive Age

Dr. Lyndon Nixon

ReTV Project Coordinator

info@retv-project.eu

@ReTV_EU

@ReTVeu

ReTV Project

retv_project