



POLIMI RECSYS CHALLENGE 2022/23



Arturo Benedetti - 10747545



Samuele Peri - 10671743



POLITECNICO
MILANO 1863



Digital



CHALLENGE OVERVIEW

The application domain is **TV shows recommendation**, the main goal of the competition is to discover which items (TV shows) a user will interact with.

The evaluation metric for this competition is **MAP@10**.

Interactions

1.8M

Number of Users

41K

Number of Items

27K

DATA ANALYSIS

Interactions from the dataset are labelled as zeros or ones:

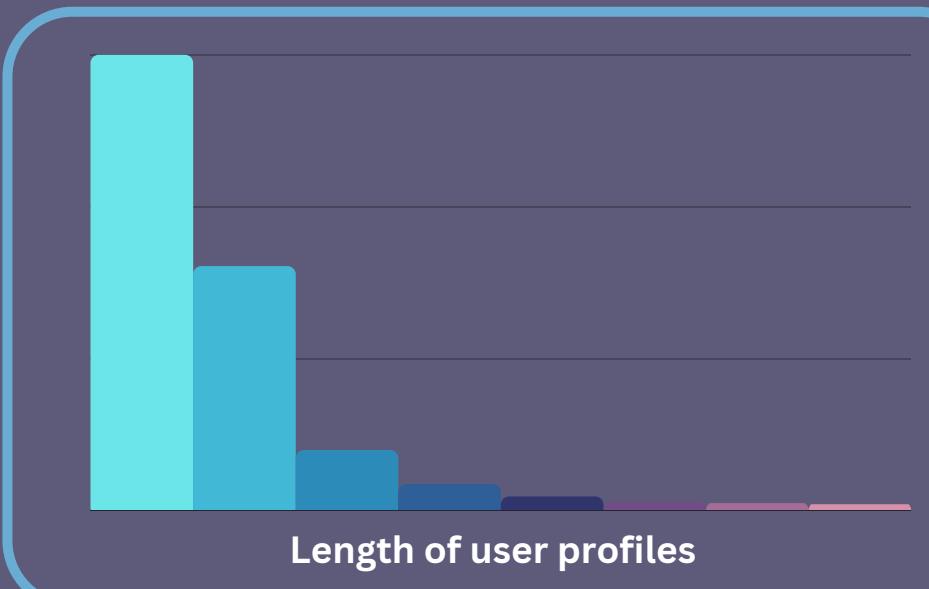
- 1 for items the user has opened the Detail Page
- 0 for the items the user has actually watched.

Interactions can be repeated, in the case of 1s a repeated interaction could mean the user has watched more than one episode.

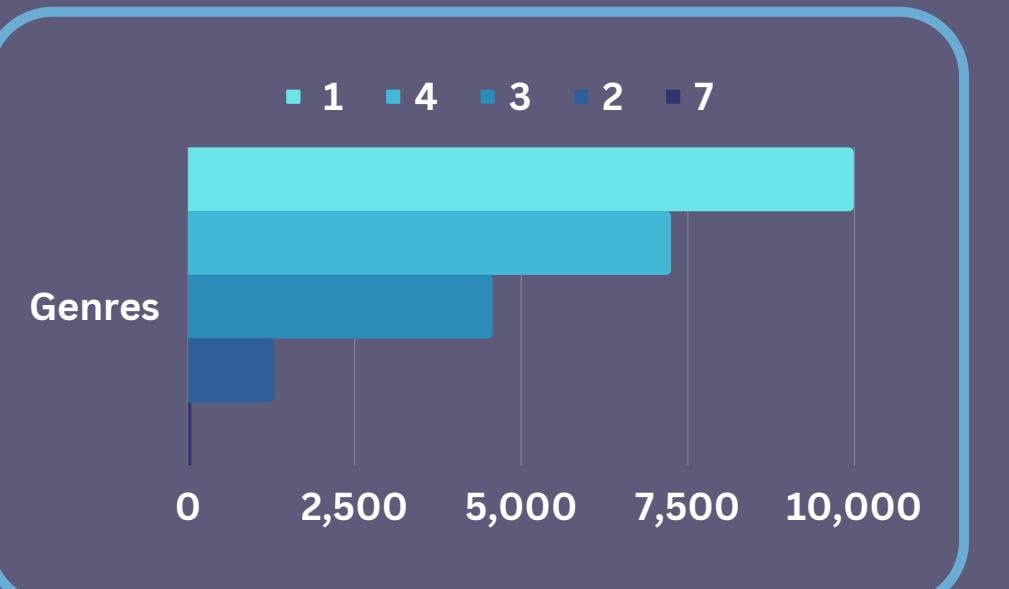
61%

61% of interactions are of type 0

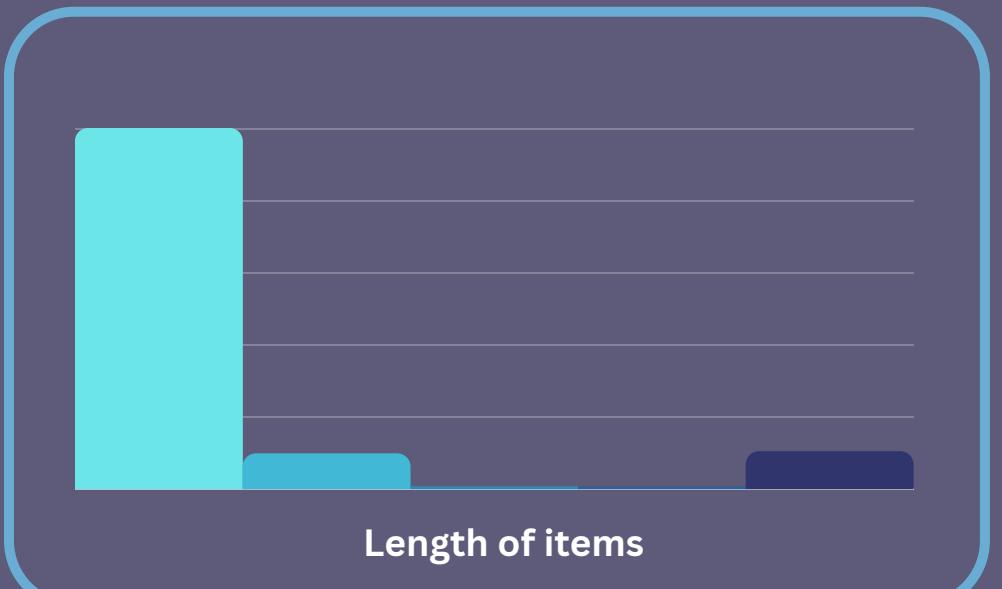
The distribution of interactions on user's profiles looks quite realistic (except some outliers): most of the users have under 80 interactions while some can have up to 8000.



Items come with information about their type (genre). Most items are of type 1, while only 5 items are of type 7.



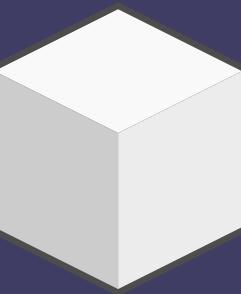
Length of items is also available, but it seems to be quite unreliable, since a lot of items have length of 1.



ROADMAP

Partial model optimization

First run of optimization on all models to get a general idea of the performances



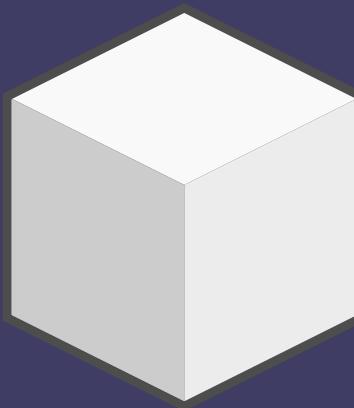
ROADMAP

Partial model optimization

First run of optimization on all models to get a general idea of the performances

Standalone SLIM Elastic

First submission with partial optimization



MAP@10 - 0.05873

ROADMAP

Partial model optimization

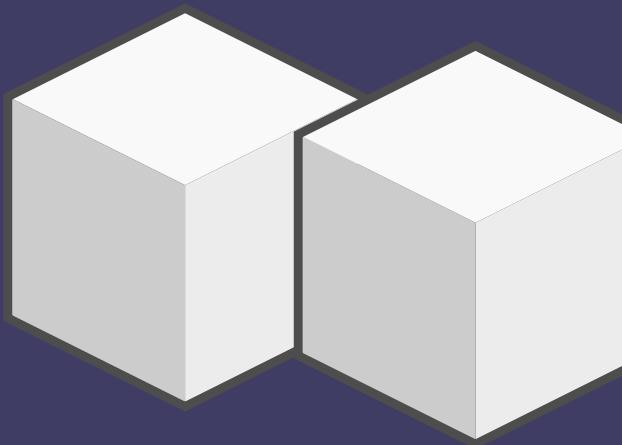
First run of optimization on all models to get a general idea of the performances

Standalone SLIM Elastic

First submission with partial optimization

Hybrid Elastic + RP3Beta

Linear combination of the scores



MAP@10 - 0.05924

ROADMAP

Partial model optimization

First run of optimization on all models to get a general idea of the performances

Standalone SLIM Elastic

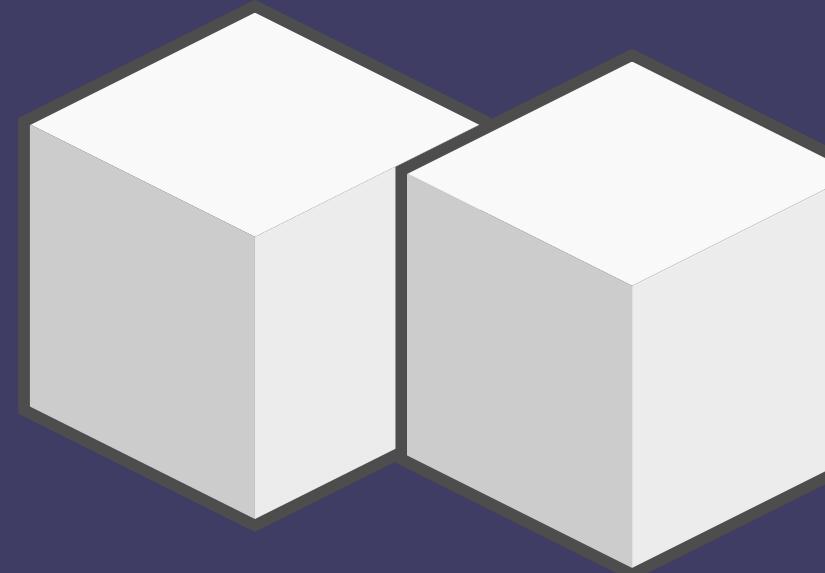
First submission with partial optimization

Hybrid Elastic + RP3Beta

Linear combination of the scores

Optimization

Baesyan search - Skopt, Optuna



MAP@10 - 0.06001

ROADMAP

Partial model optimization

First run of optimization on all models to get a general idea of the performances

Standalone SLIM Elastic

First submission with partial optimization

Hybrid Elastic + RP3Beta

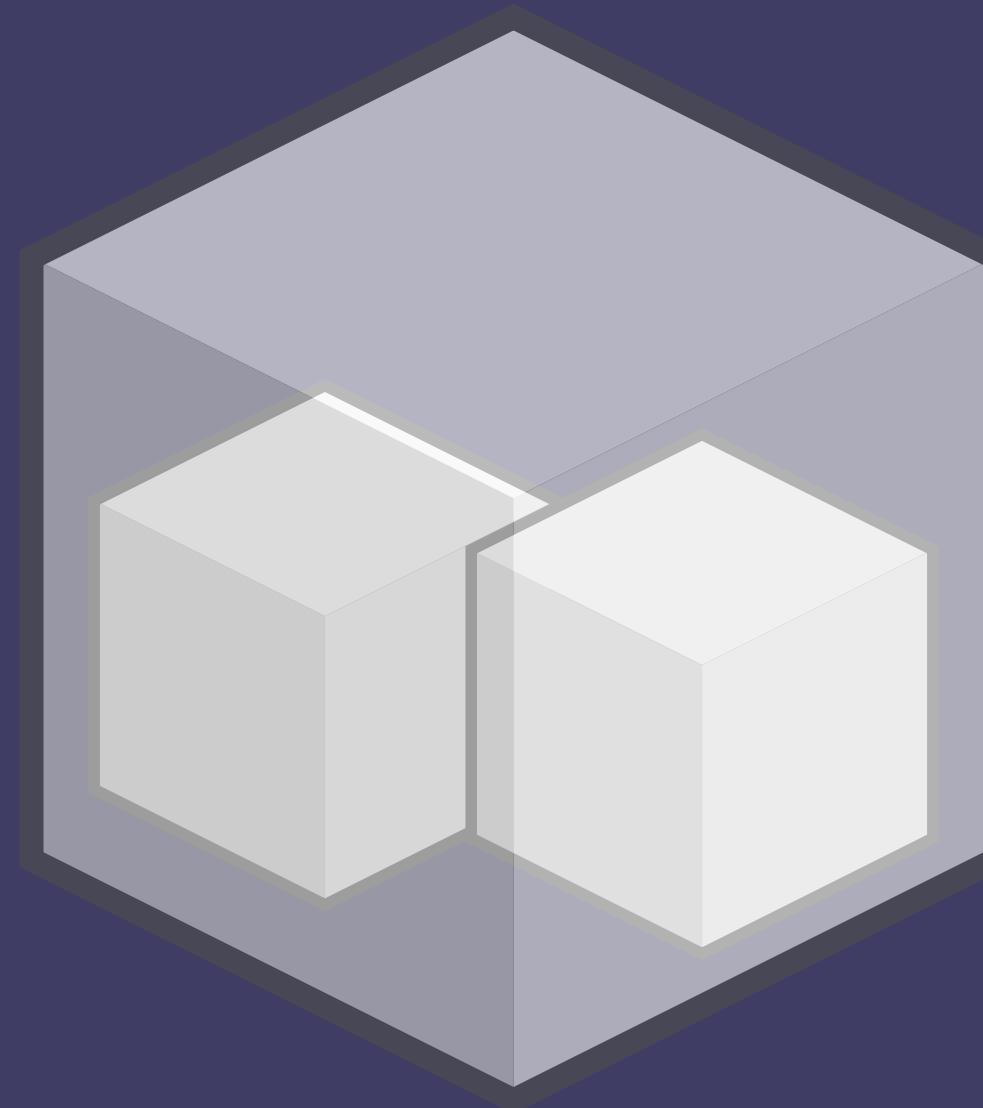
Linear combination of the scores

Optimization

Baesyan search - Skopt, Optuna

Boosting

XGBoost reranker



MAP@10 - 0.06180

ROADMAP

Partial model optimization

First run of optimization on all models to get a general idea of the performances

Standalone SLIM Elastic

First submission with partial optimization

Hybrid Elastic + RP3Beta

Linear combination of the scores

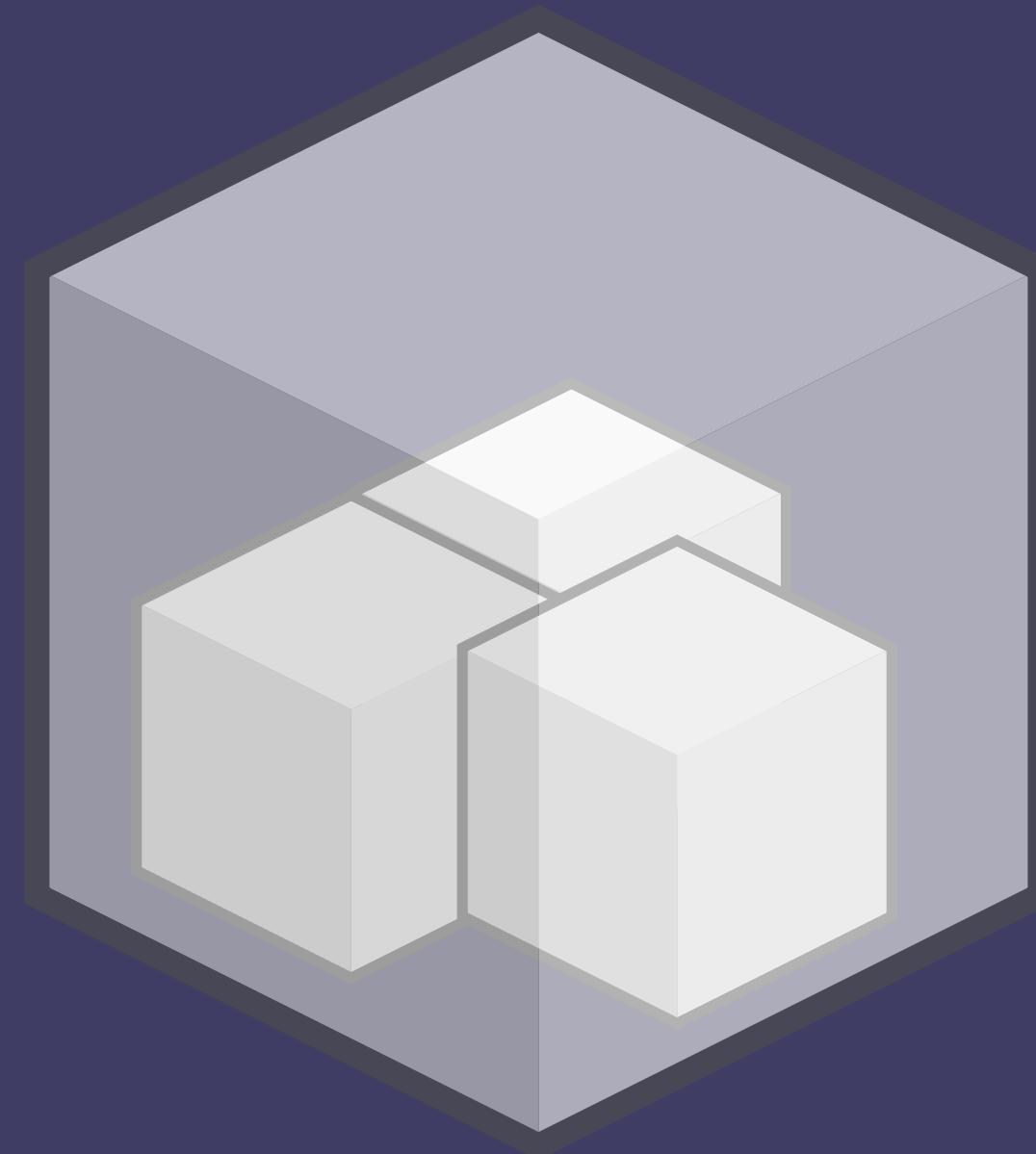
Optimization

Baesyan search - Skopt, Optuna

Boosting

XGBoost reranker

Added IALS to Hybrid



MAP@10 - 0.06205

ROADMAP

Partial model optimization

First run of optimization on all models to get a general idea of the performances

Standalone SLIM Elastic

First submission with partial optimization

Hybrid Elastic + RP3Beta

Linear combination of the scores

Optimization

Baesyan search - Skopt, Optuna

Boosting

XGBoost reranker

Added IALS to Hybrid

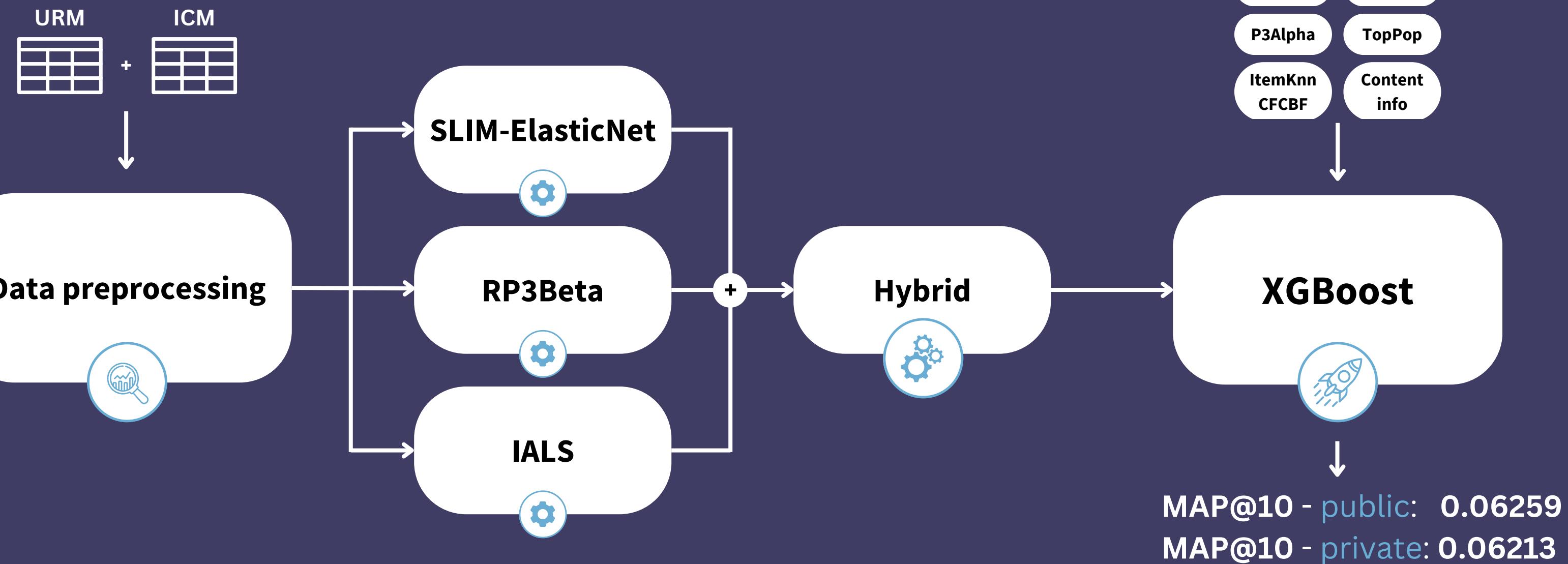
Data preprocessing

ICM Stacking + Repeated Interactions bonus

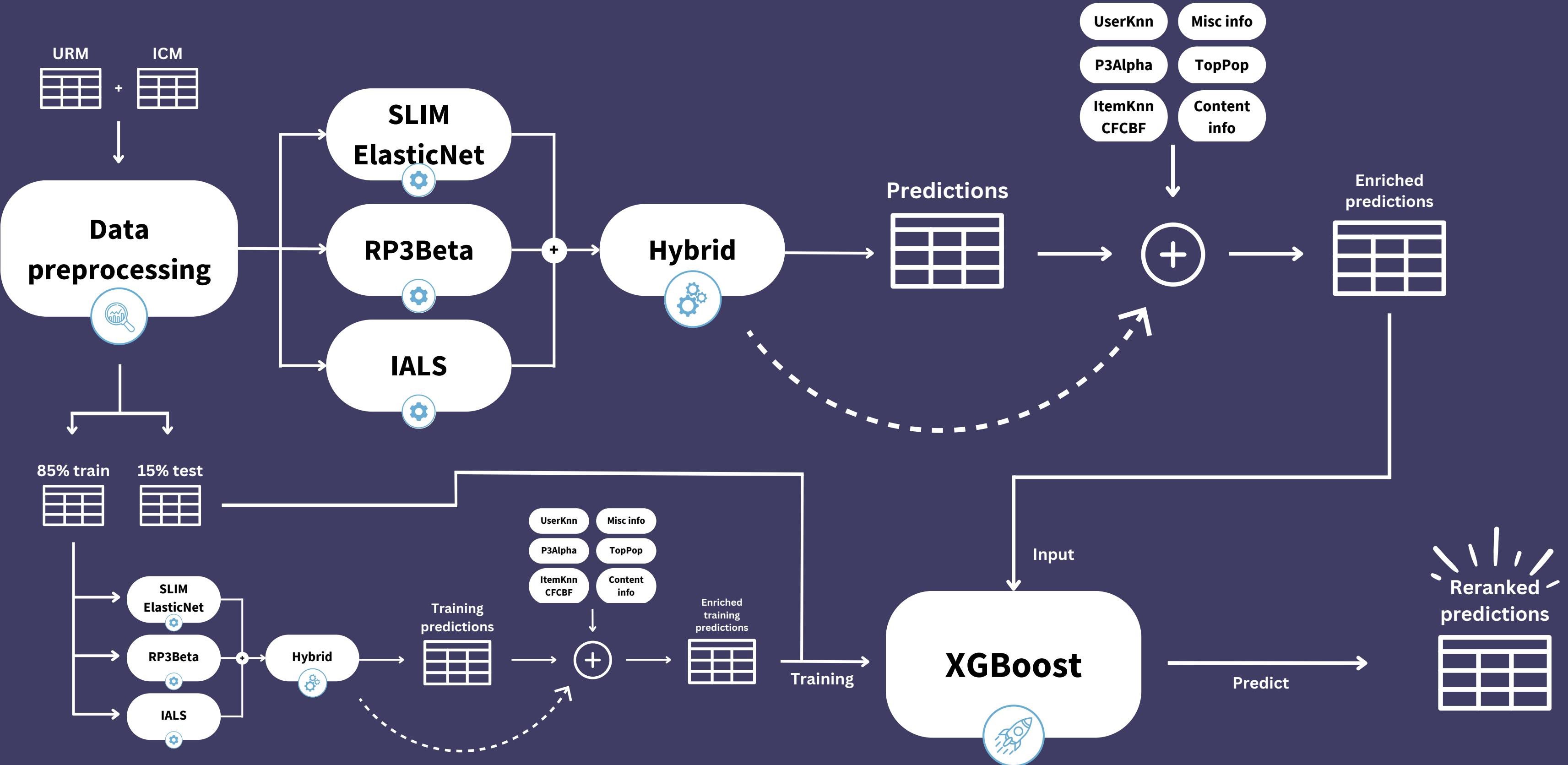


MAP@10 - 0.06259

OUR SOLUTION



ARCHITECTURE





~20% of items
changed after reranking

Features:

- Item popularity
- Item length
- Item genre
- User profile length
- Scores of other models

Cutoff: 35

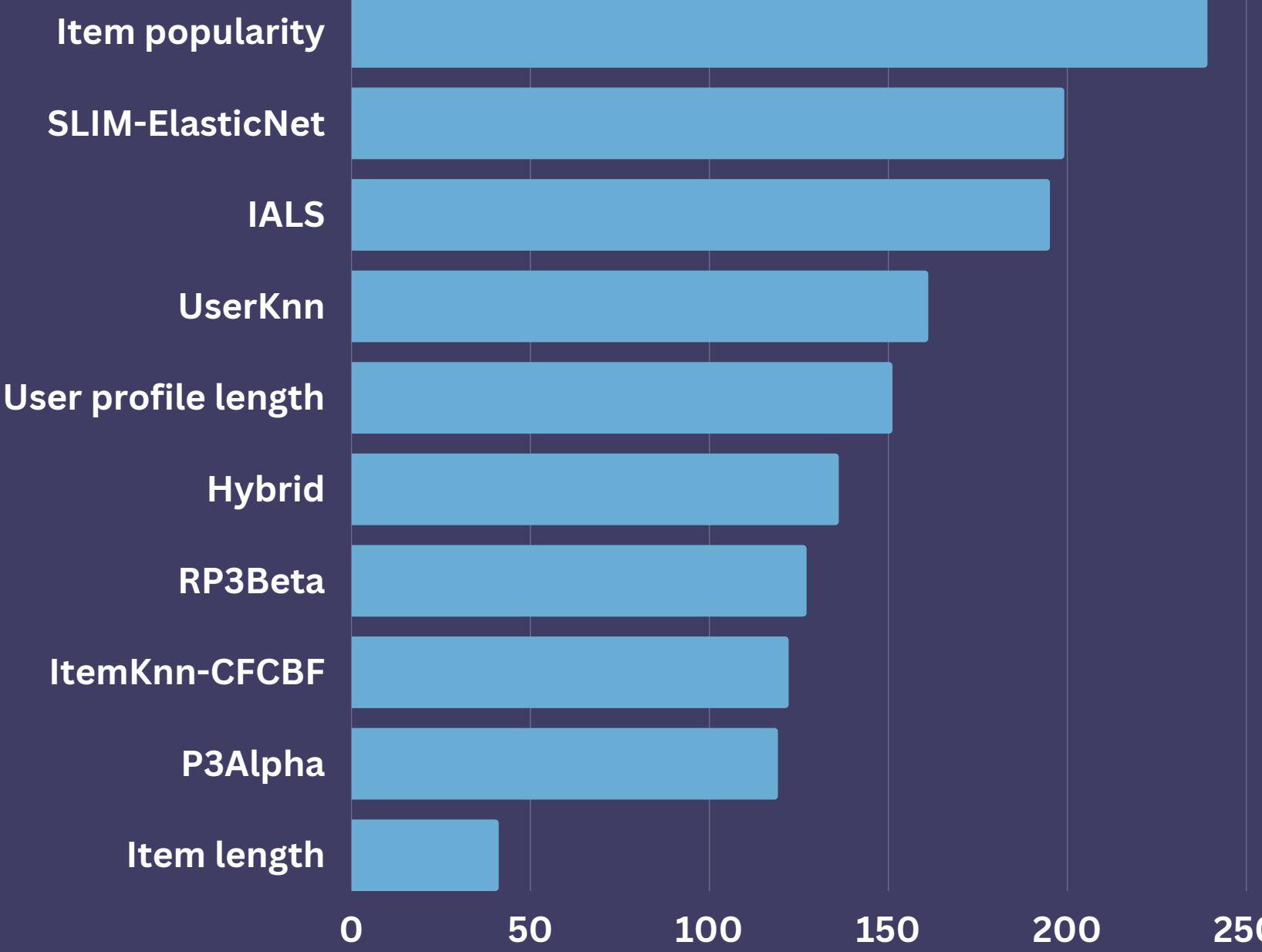
Local scores

Hybrid: 0.0458

Boosted: **0.0470**

*custom evaluator

Feature importance

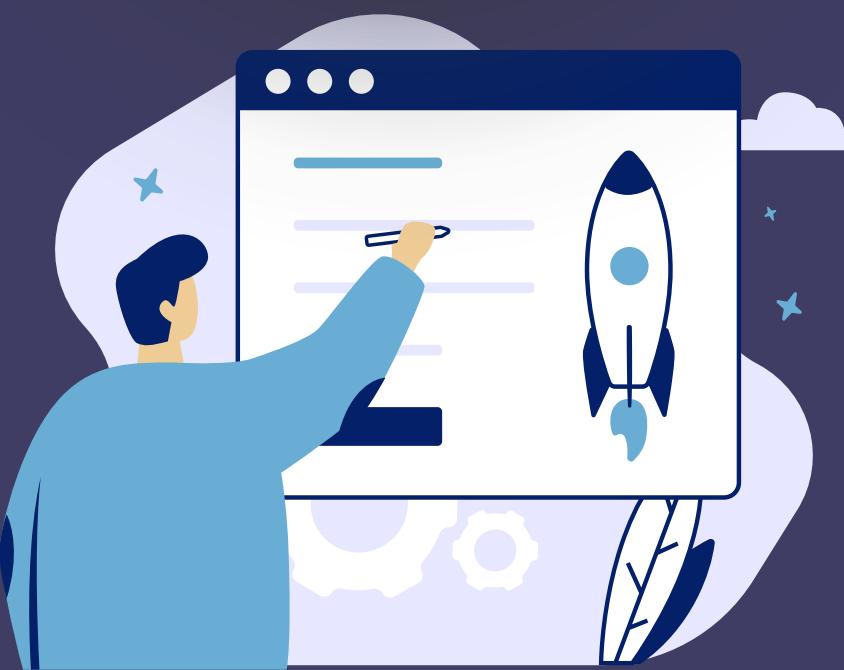


WHAT DIDN'T WORK

Removing
Outliers

Stacking

User-tailored
recs



IMPROVEMENTS



Stacking



Impression analysis



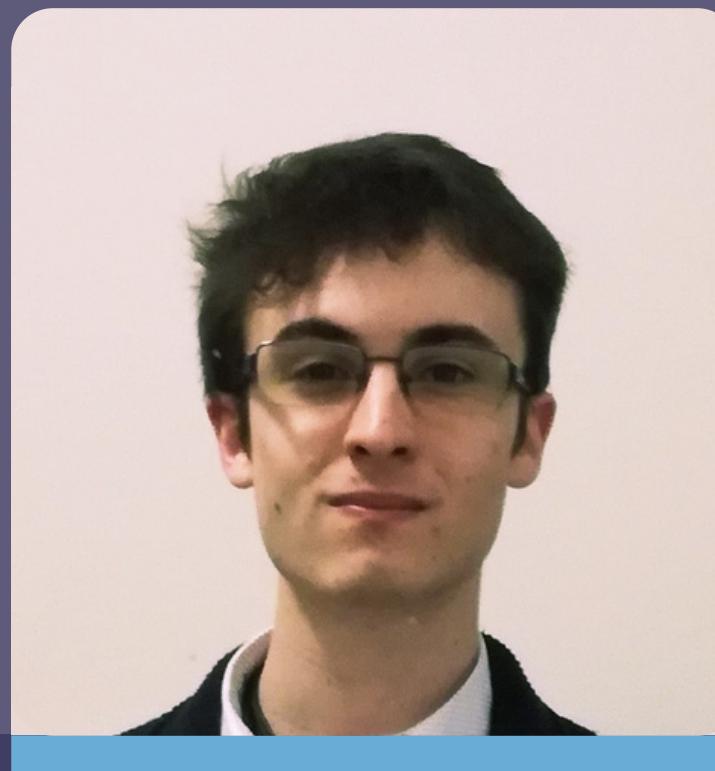
Improve preprocessing



Noise reduction



TEAM



Arturo Benedetti

Human-Computer Interaction & Design
@EIT Digital Master School
(POLIMI/ UPS)



Benedart



Samuele Peri

Data Science
@EIT Digital Master School
(POLIMI/ KTH)



john-galt-10