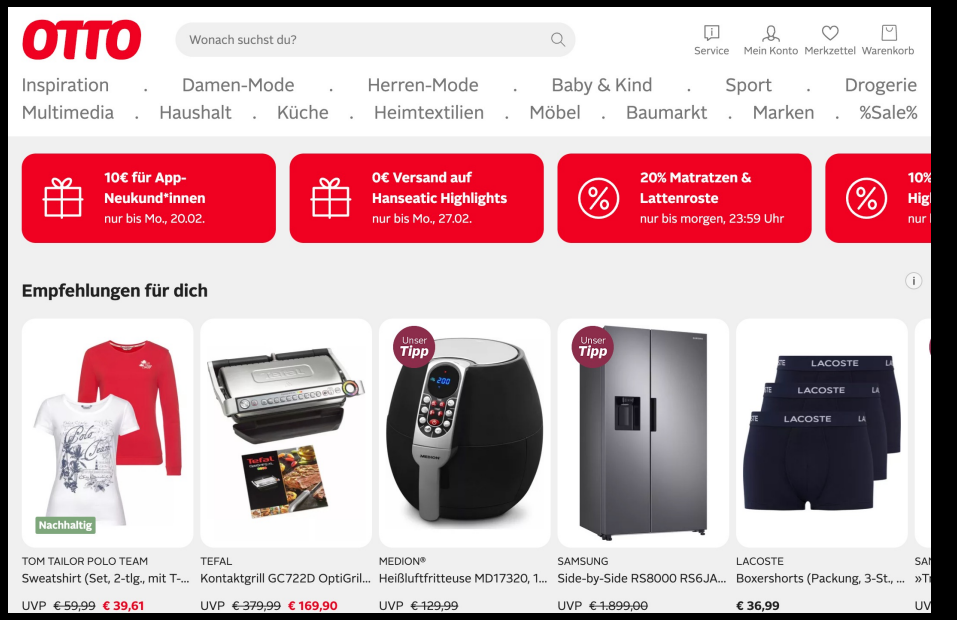
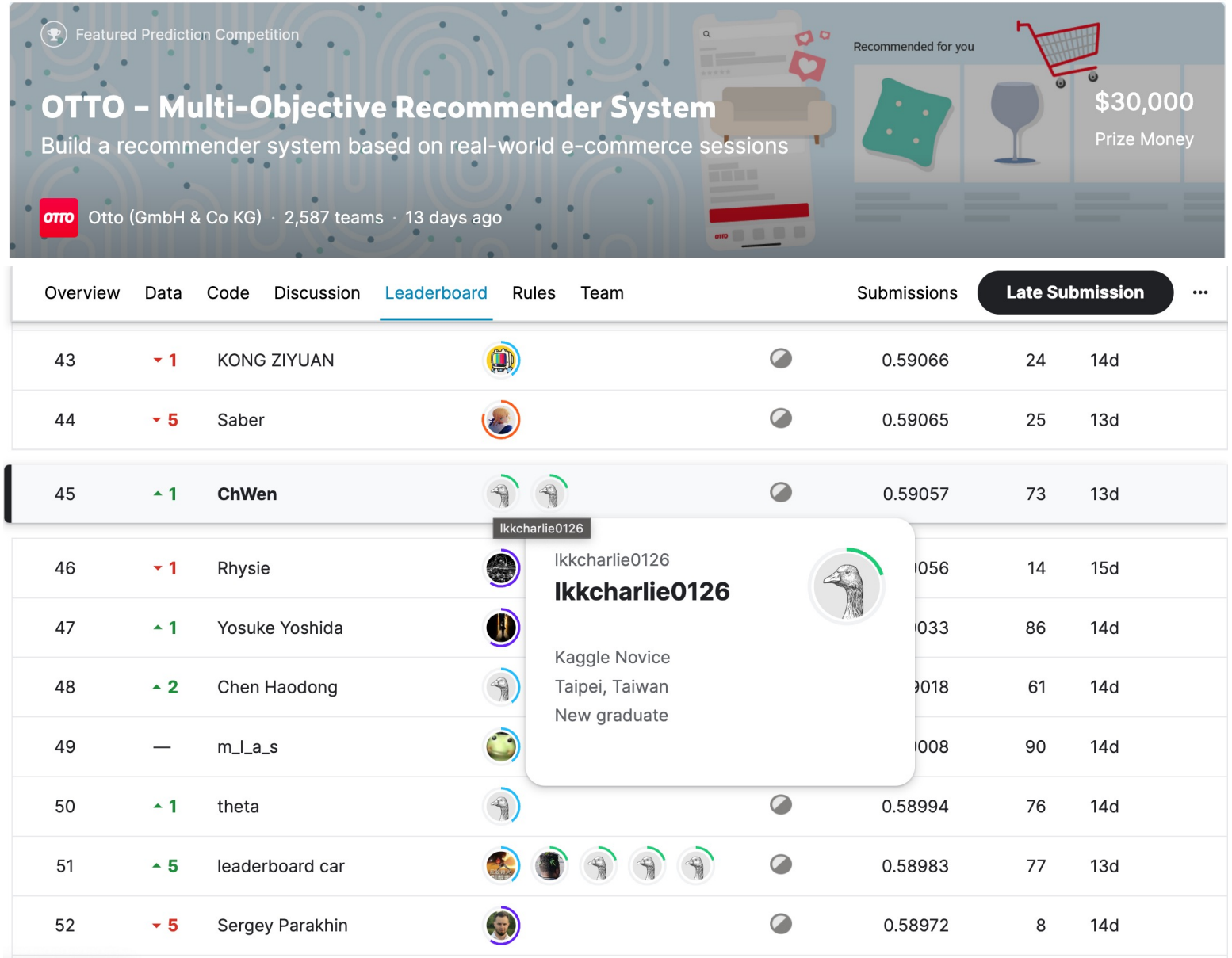
Kaggle – OTTO Multi-Objective Recommender System

Silver medal
Rank top 2%
(45 / 2587)





Dataset Description

```
{
    "session": 42,
    "events": [
        { "aid": 0, "ts": 1661200010000, "type": "clicks" },
        { "aid": 1, "ts": 1661200020000, "type": "clicks" },
        { "aid": 2, "ts": 1661200030000, "type": "clicks" },
        { "aid": 2, "ts": 1661200040000, "type": "carts"  },
        { "aid": 3, "ts": 1661200050000, "type": "clicks" },
        { "aid": 3, "ts": 1661200060000, "type": "carts"  },
        { "aid": 4, "ts": 1661200070000, "type": "clicks" },
        { "aid": 2, "ts": 1661200080000, "type": "orders" },
        { "aid": 3, "ts": 1661200080000, "type": "orders" }
    ]
}
```

(Session = User, Aid = Item)

Dataset	#sessions	#items	#events	#clicks	#carts	#orders
Train (week 1~4)	12,899,779	1,855,603	216,716,096	194,720,954	16,896,191	5,098,951
Test (week 5)	1,671,803	1,019,357	13,851,293	12,340,303	1,155,698	355,292

Cold Start

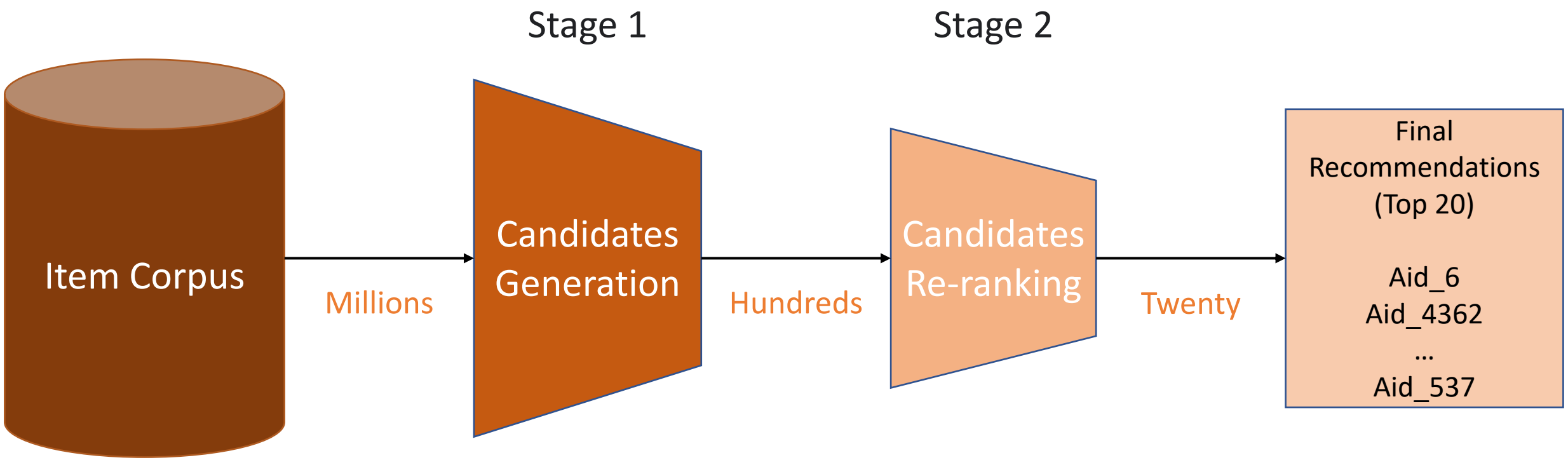
Evaluation Metric

Weighted Recall@20

$$R_{type} = \frac{\sum_{i}^{N} |\{\text{predicted aids}\}_{i,type} \cap \{\text{ground truth aids}\}_{i,type}|}{\sum_{i}^{N} \min(20, |\{\text{ground truth aids}\}_{i,type}|)}$$

$$score = 0.10 \cdot R_{clicks} + 0.30 \cdot R_{carts} + 0.60 \cdot R_{orders}$$

Flow Chart

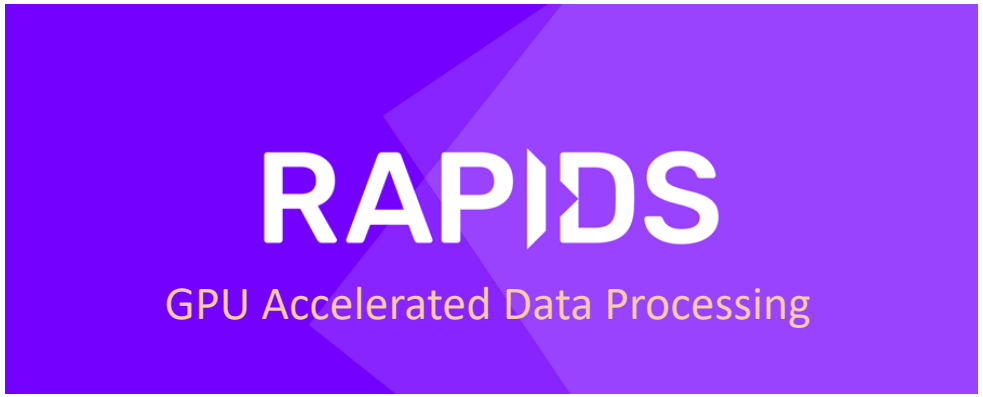

Stage 1
Stage 2
Item Corpus
Millions
Candidates Generation
Hundreds
Candidates Re-ranking
Twenty
Final Recommendations (Top 20)
Aid_6
Aid_4362
...
Aid_537

Candidates Generation

Generate 100 item candidates for each user based on following logics:

- Visited aids in session
  (User's history events)

- Co-visitation matrix
  - Click2Click (in 5mins) (time-weighted)
  - Click2Cart (in 30 mins) (time-weighted)
  - Click2Order (in 60 mins) (time-weighted)
  - Any2Click (in 1 day) (time-weighted)
  - Any2Buy (in 1 day) (type-weighted)
  - Buy2Buy (in 14 days)



25x faster!

- Item-item Word2Vec similarity
  - (weighted by w2v similarity with the last aid in session)

- Popular items in the last week

Features Generation

Generate 80+ features by only user_id, item_id, type, and timestamp

- **User-based (14)**
  - number of events (1)
  - number of each event (3)
  - number of unique aids (1)
  - buy ratio of user (1)
  - type statistics of user (2)
  - time statistics of user (4)
  - time spend on each item (2)
- **Item-based (28) (train + test)**
  - count of (any/click/cart/order) (4)
  - number of unique users (1)
  - bought ratio of item (1)
  - type statistics of item (2)
  - time statistics of item (4)
  - user staying time (2)
- **User-Item interaction (4)**
  - number of item (clicked/carted/ordered) by user (3)
  - Ranking of rule-based ranker (1)
- **Collaborative filtering similarity (5)**
  - average CF cosine similarity with all aids (1)
  - weighted average CF cosine similarity with all aids (1)
  - CF cosine similarity with last 3 aids (3)
- **Word2Vec (14)**
  - w2v embedding (8)
  - average w2v cosine similarity with all aids (1)
  - weighted average w2v cosine similarity with all aids (1)
  - w2v cosine similarity with last 4 aids (4)
- **Co-visitation score (18)**
  - weighted average co-visitation score with all aids (1 * 6)
  - co-visitation score with last 2 aids (2 * 6)

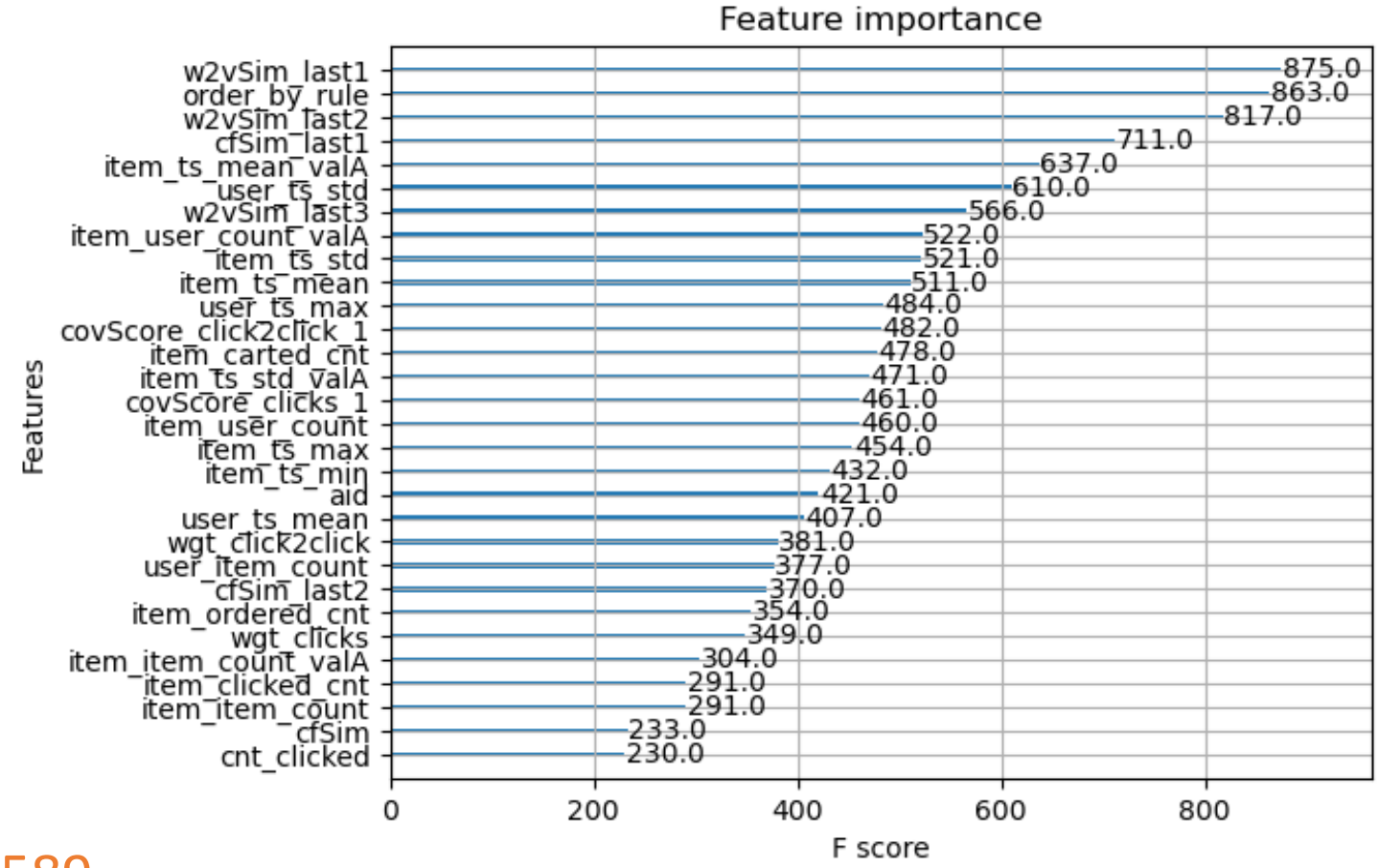
Candidates Re-ranking

Ranker: XGBoost ranker

Negative sampling rate: 4%

Loss: Pairwise

Regularization: L1 & L2



Single XGBoost ranker: LB 0.589

Ensemble of 5 XGBoost rankers with different hyperparameters : LB 0.590

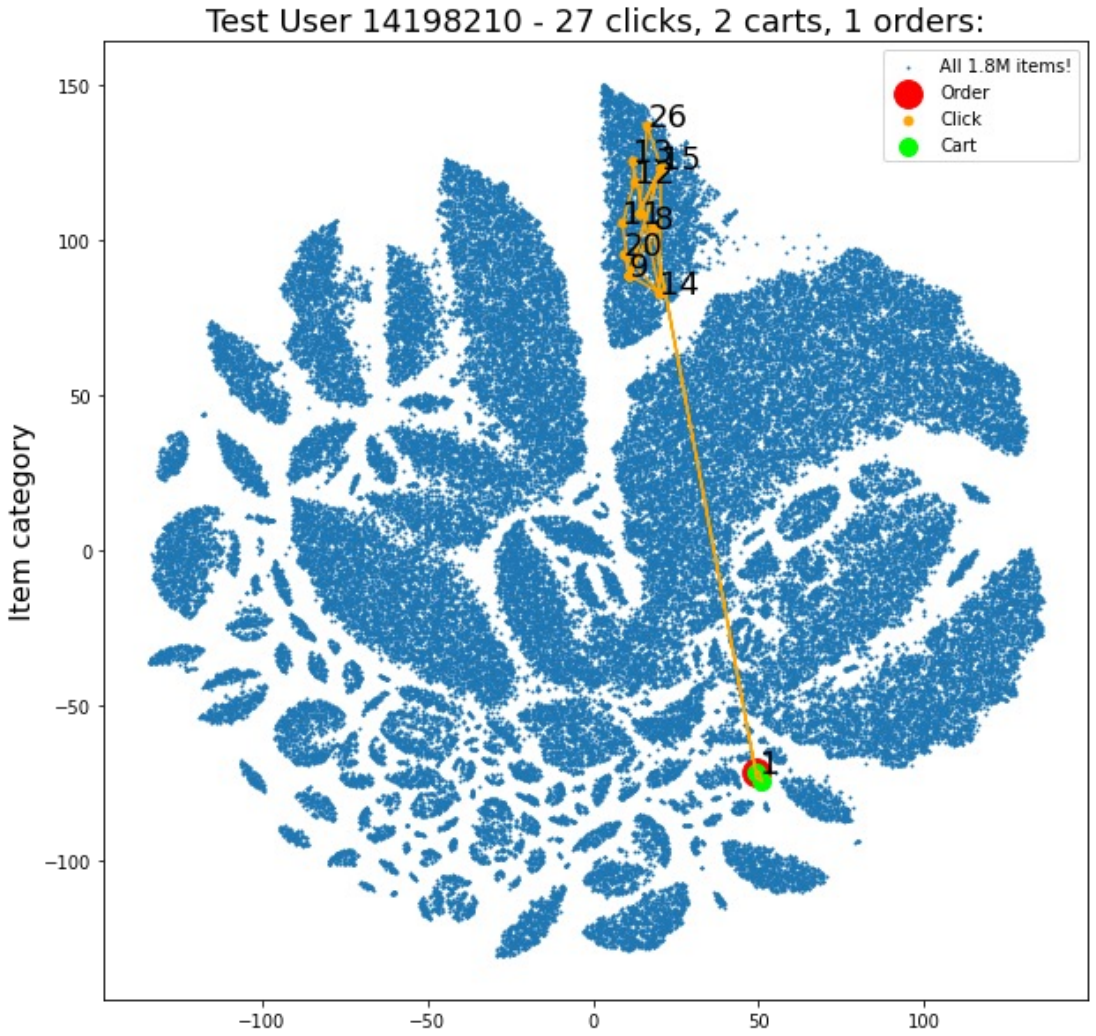
What improved:

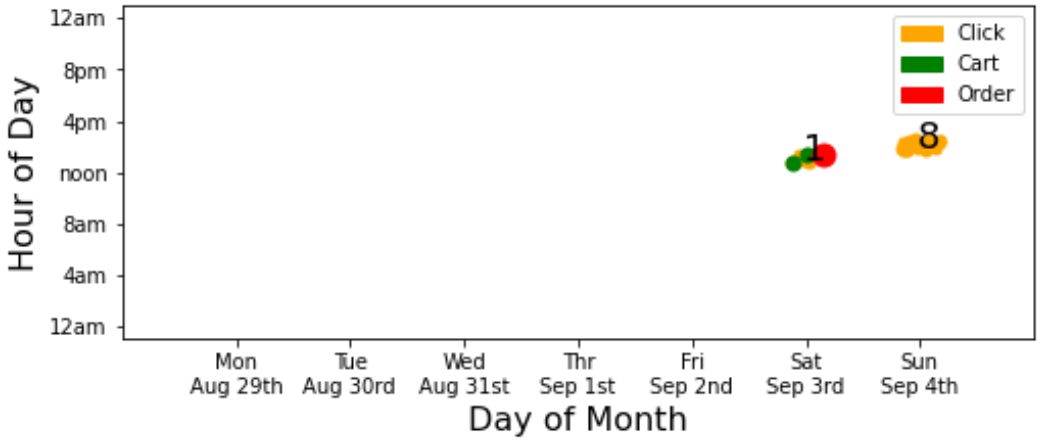
- **Rule-based only**: 0.5749
- **XGB re-ranker**: 0.5776
- **[Prevent XGB over-fitting]** not shared co-visitation mat: 0.5784
- **[Tune rule-based]** Tune co-visitation mat hyperparameter: 0.5807
- **[Add XGB feature]** Add co-visitation score with all aids: 0.5829
- **[Tune XGB]** Tune XGB hyperparameter: 0.5834
- **[Add XGB feature]** w2v similarity with all aids: 0.5835
- **[New rule-based logic]** More co-visitation mat : 0.5849
- **[More candidates]** 50 -> 100 candidates: 0.5854
- **[Add XGB feature]** CF similarity: 0.5865
- **[Add XGB feature]** time weighted CF & w2v similarity: 0.5867
- **[Add XGB feature]** CF & w2v similarity & co-visitation score with last aids: 0.5893
- **[New rule-based logic]** consider w2v similarity in rule-based: 0.5903
- **[Ensemble]** 5 XGB models: 0.5906

What didn't work:

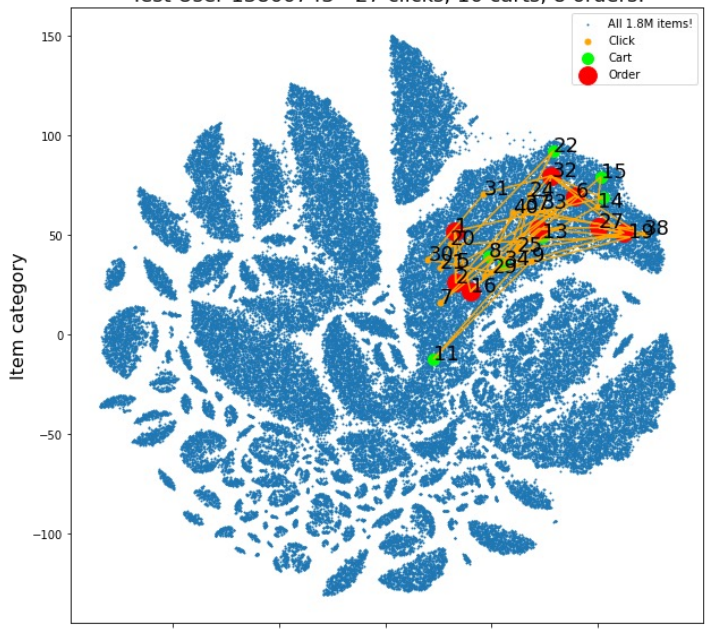
- **feature selection**
- **100 -> 200 candidates**
- **w2v embedding as features**

Visualize of User Behavior & Item Matrix Factorization by TSNE

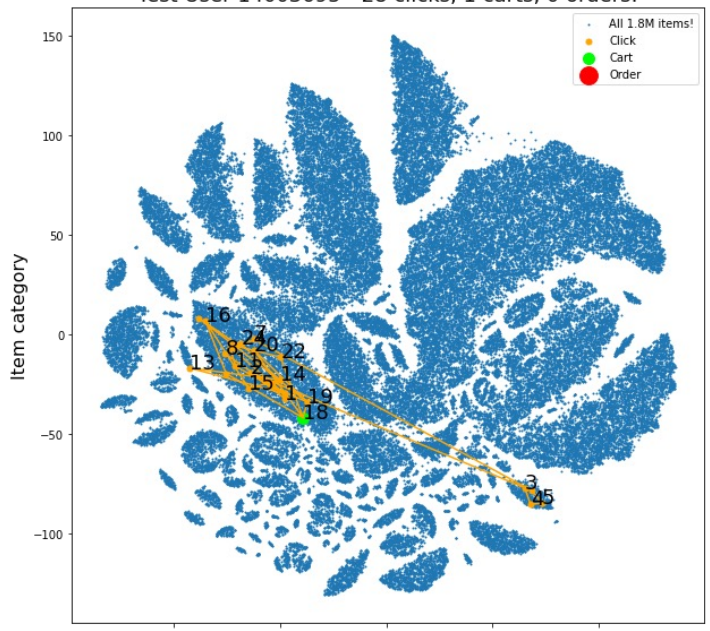




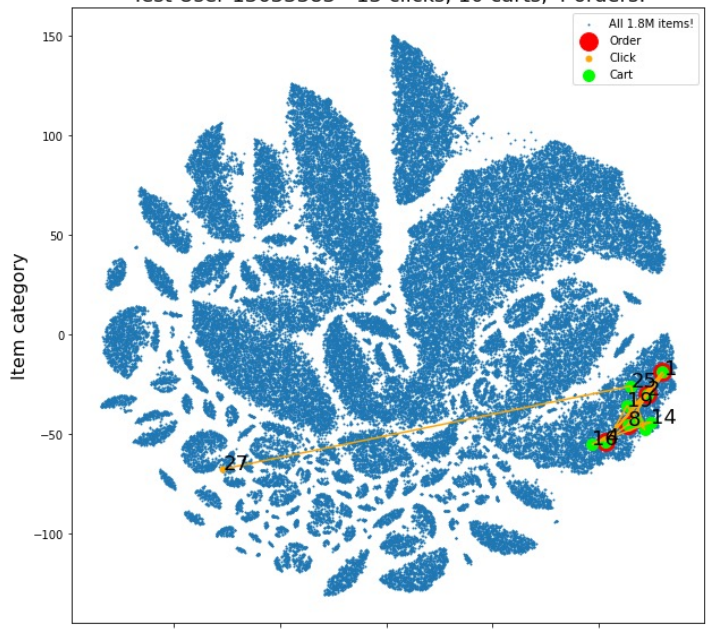
Test User 13866743 - 27 clicks, 16 carts, 8 orders:



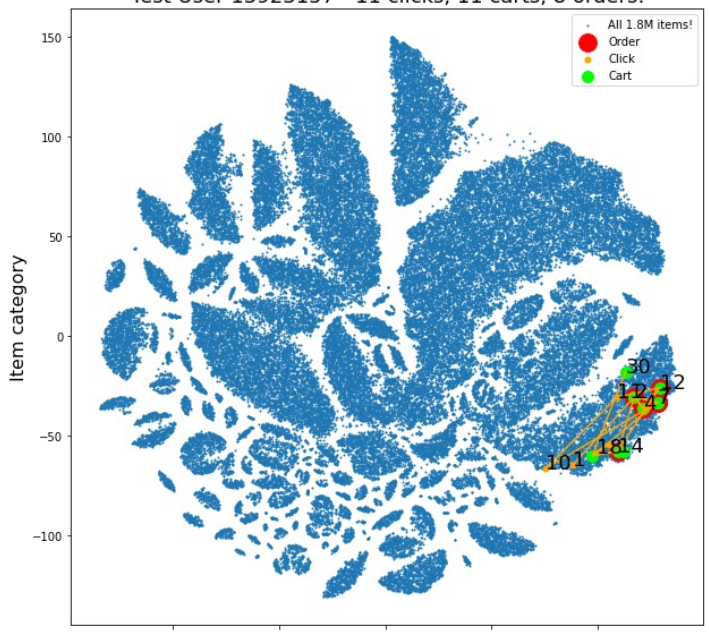
Test User 14003695 - 28 clicks, 1 carts, 0 orders:



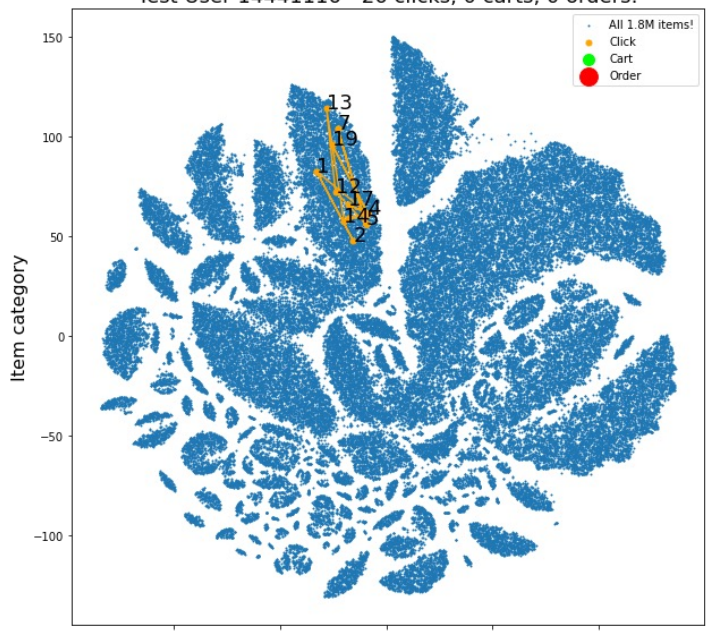
Test User 13633383 - 13 clicks, 10 carts, 4 orders:



Test User 13923157 - 11 clicks, 11 carts, 8 orders:



Test User 14441116 - 26 clicks, 0 carts, 0 orders:



Test User 12975893 - 58 clicks, 14 carts, 0 orders: