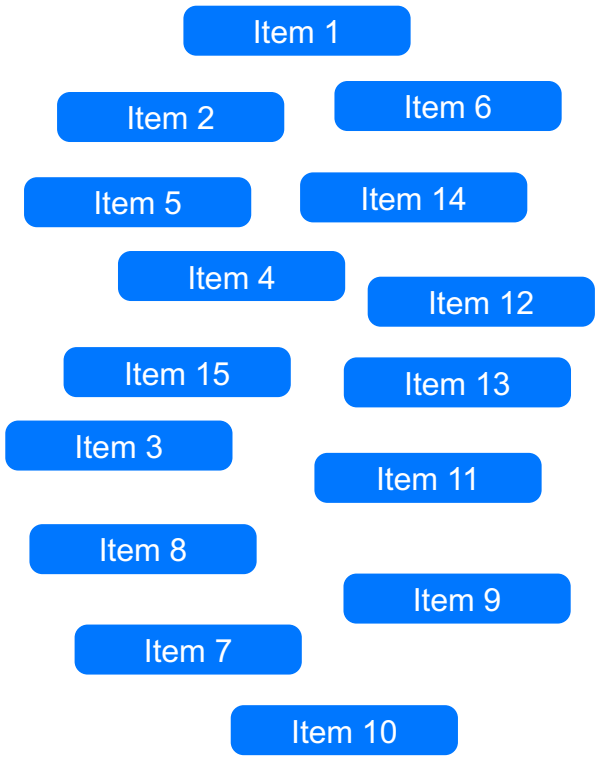
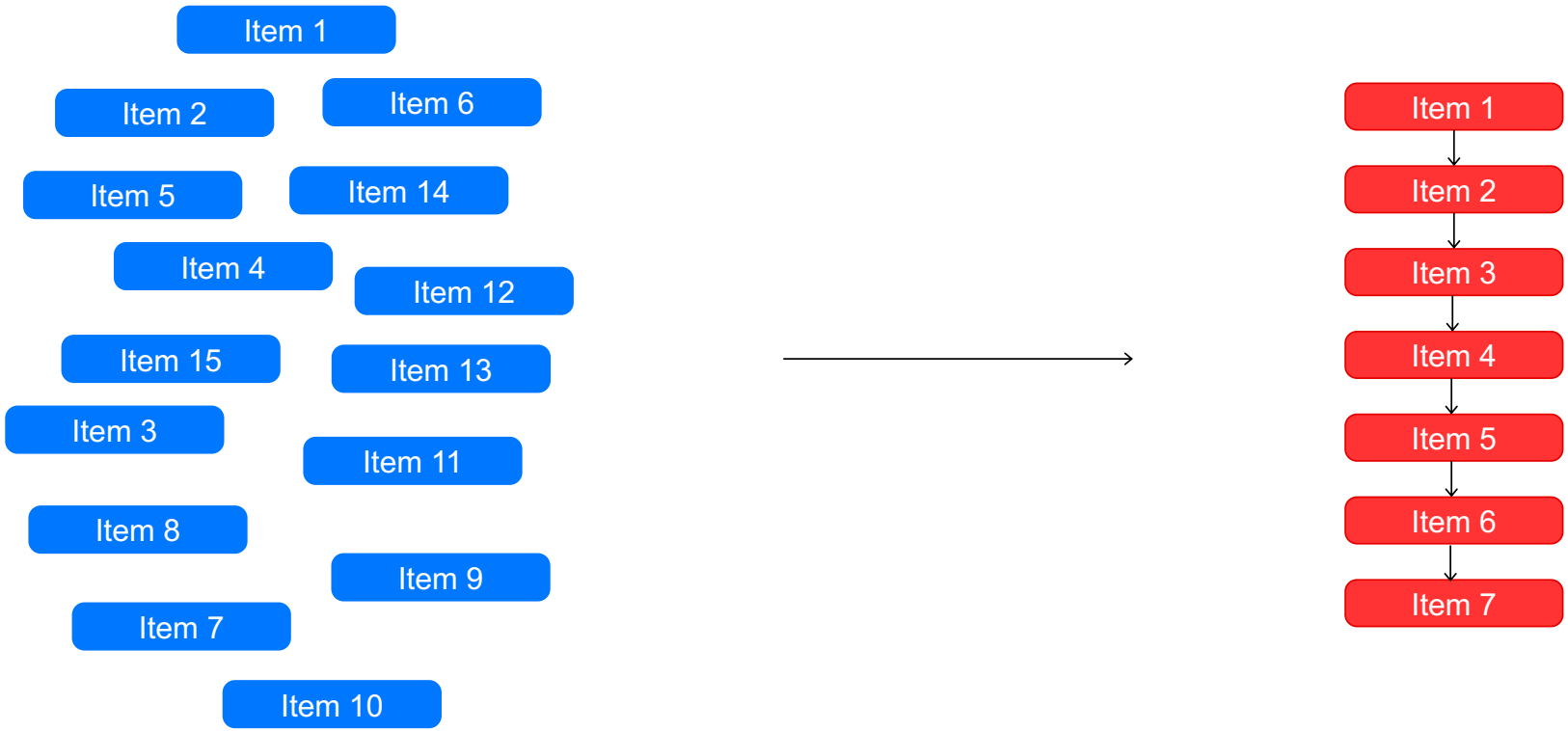
Recommendation systems: from theory to practice

Danila Chiryatnikov, Yandex.Q


Item 1
Item 2
Item 6
Item 5
Item 14
Item 4
Item 12
Item 15
Item 13
Item 3
Item 11
Item 8
Item 9
Item 7
Item 10


Item 1
Item 2
Item 6
Item 5
Item 14
Item 4
Item 12
Item 15
Item 13
Item 3
Item 11
Item 8
Item 9
Item 7
Item 10
Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7

01

Recommendation system structure

Recommendation system

Recommendation system

Main steps

Recommendation system

Main steps

1. Candidate generation

Recommendation system

Main steps

1. Candidate generation
2. Filtering

Recommendation system

Main steps

1. Candidate generation
2. Filtering
3. Ranking

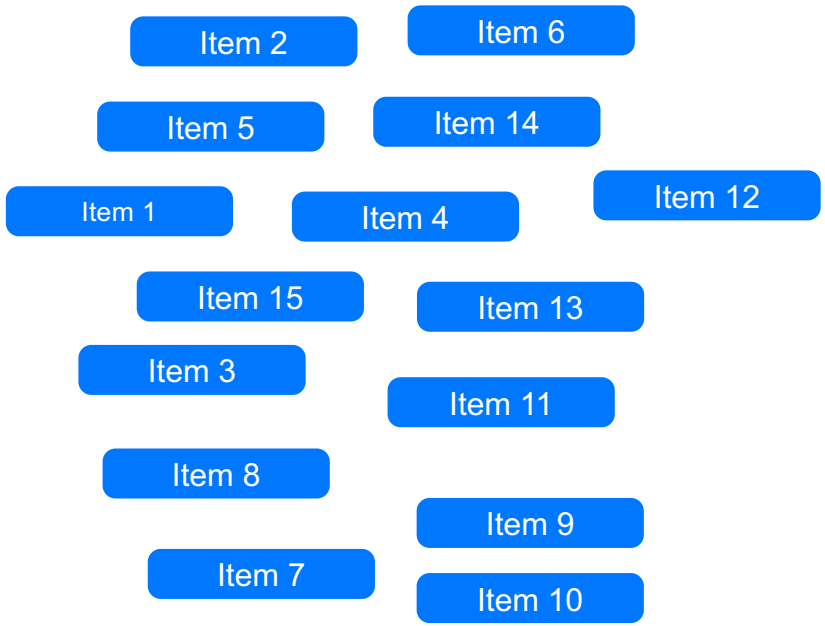
Recommendation system

Main steps

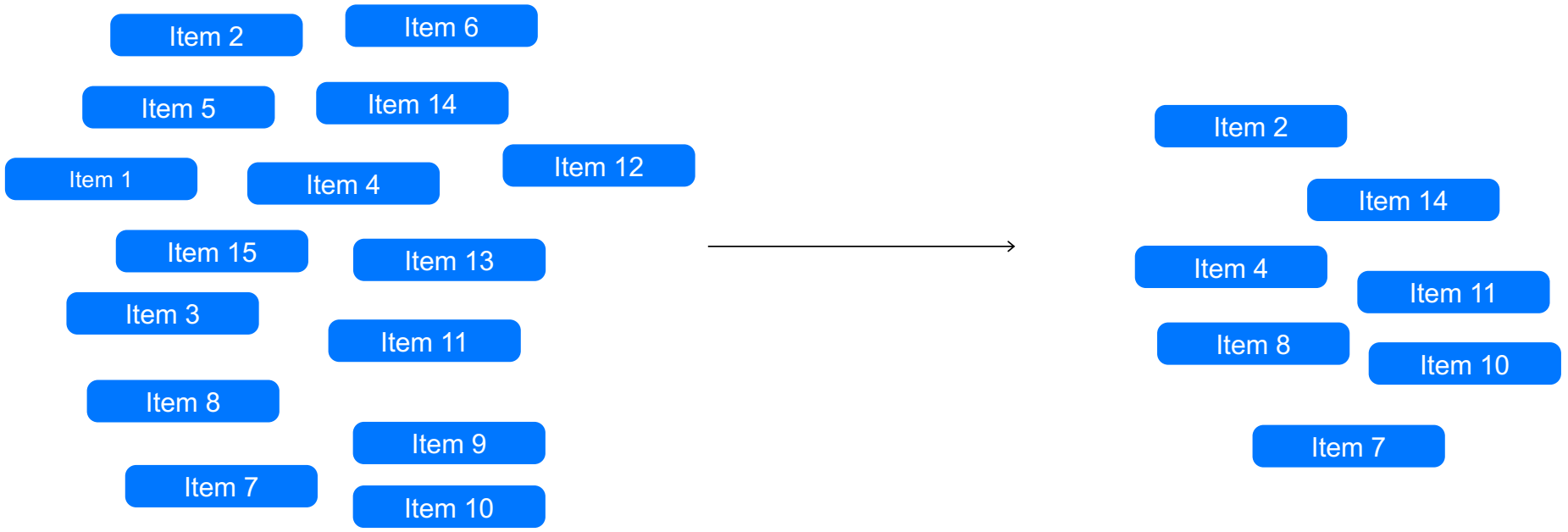
1. Candidate generation
2. Filtering
3. Ranking
4. Filtering
5. Reranking
6. ….

Candidate generation

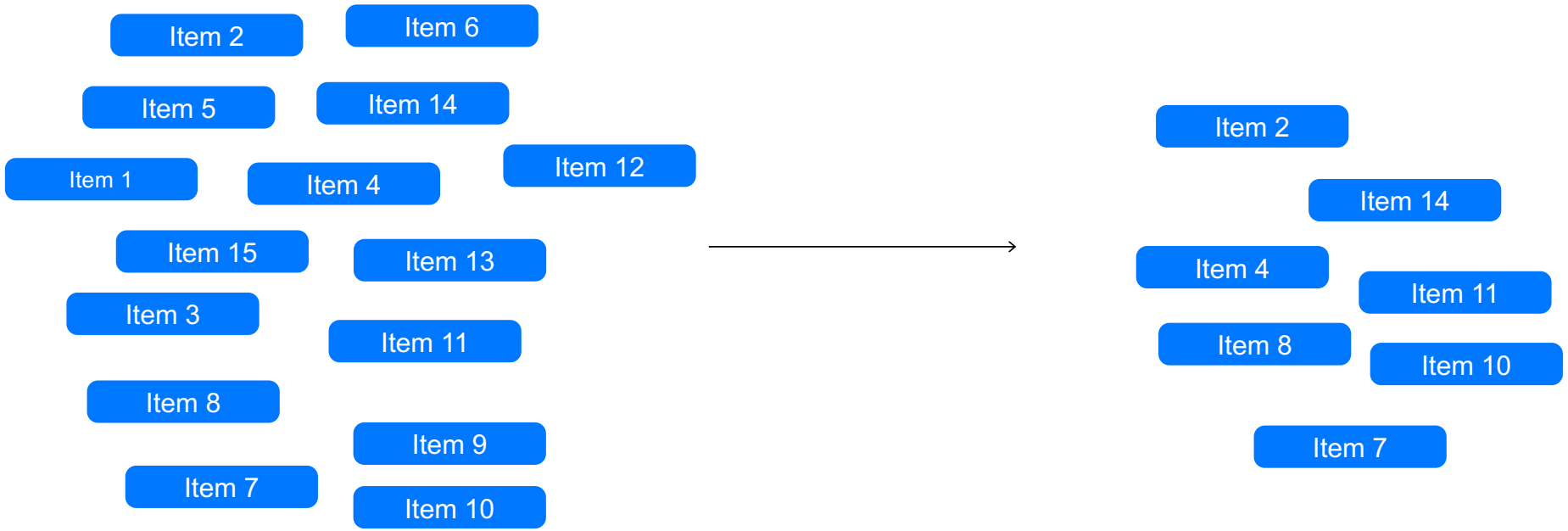
Candidate generation



Candidate generation



Candidate generation



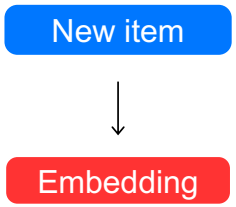
- HNSW
- Embeddings

Candidate generation

Candidate generation

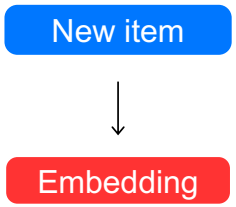
New item

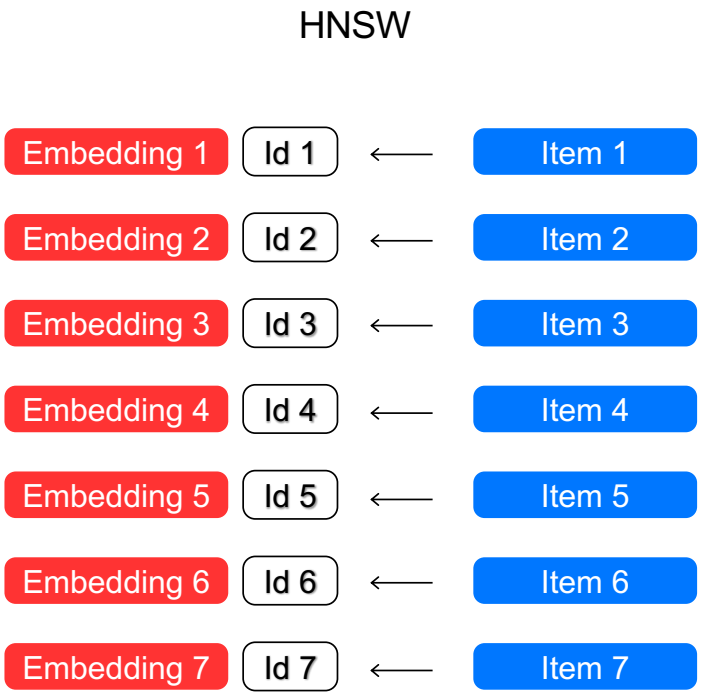
Candidate generation



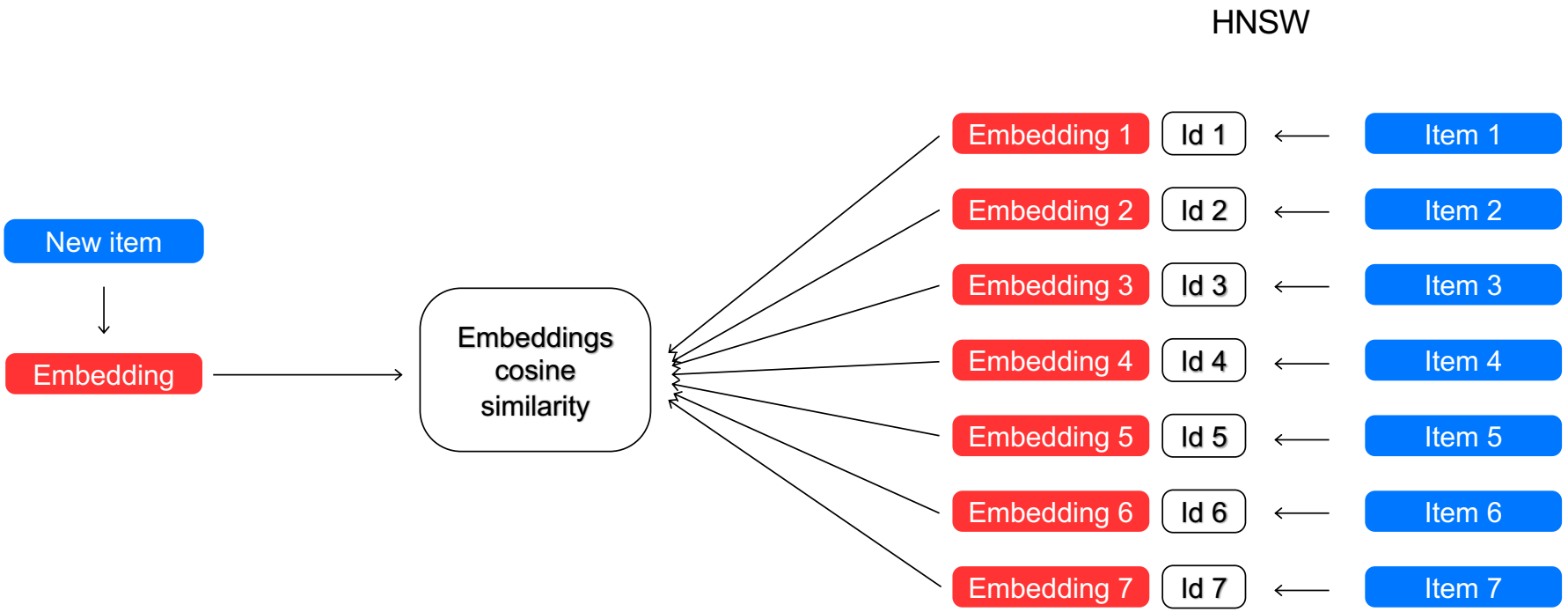
Candidate generation

Candidate generation

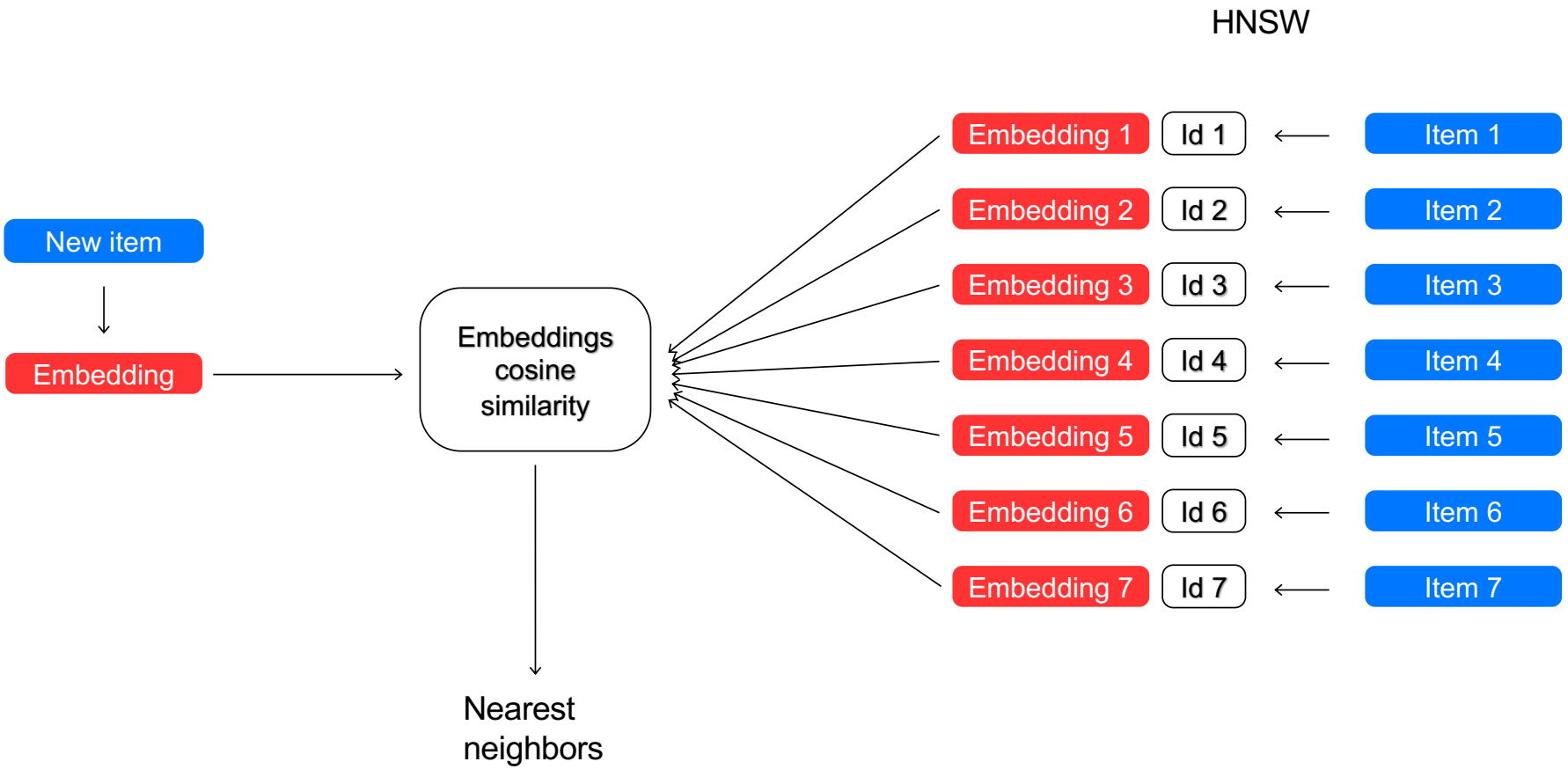




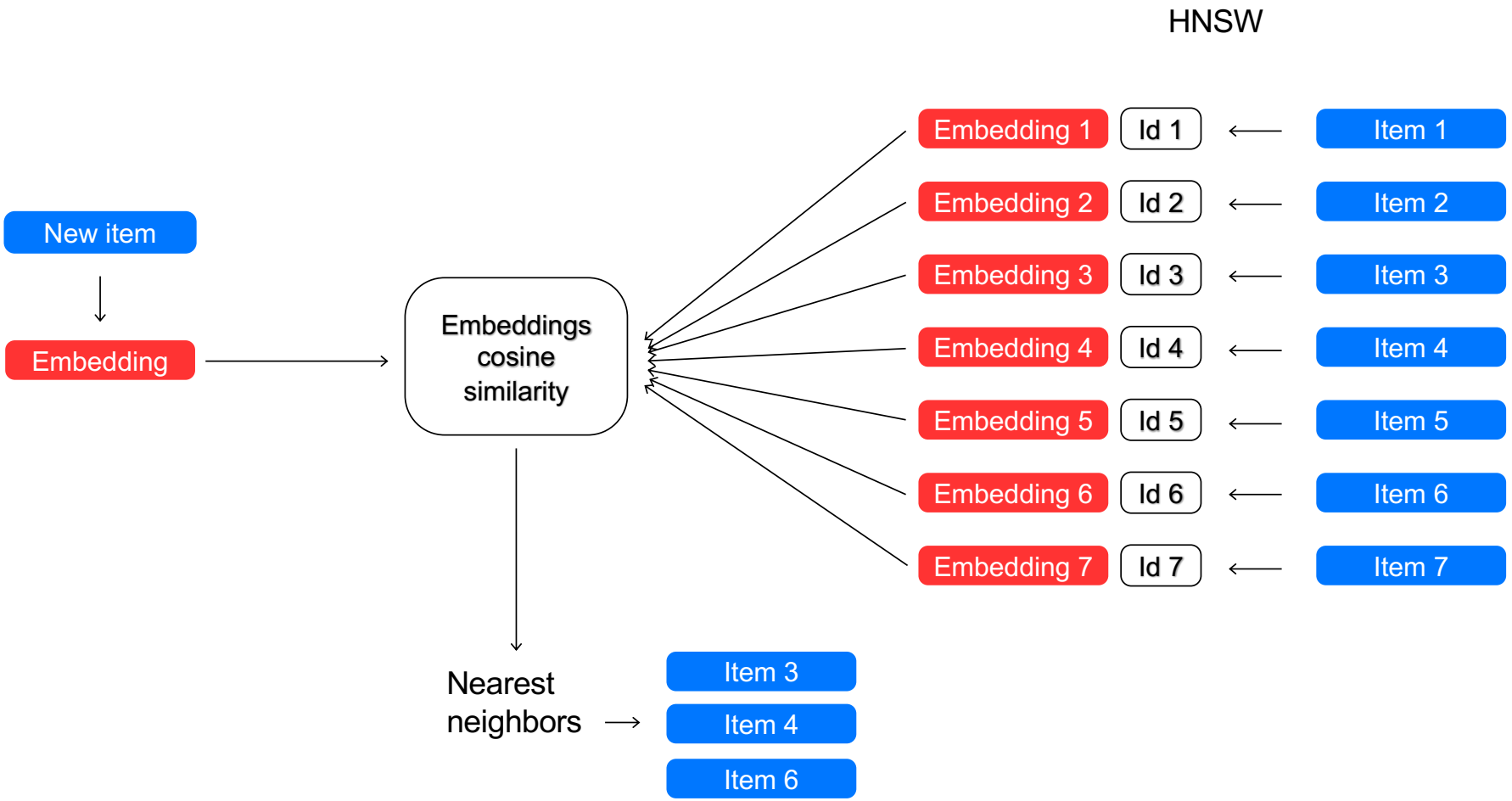
Candidate generation



Candidate generation

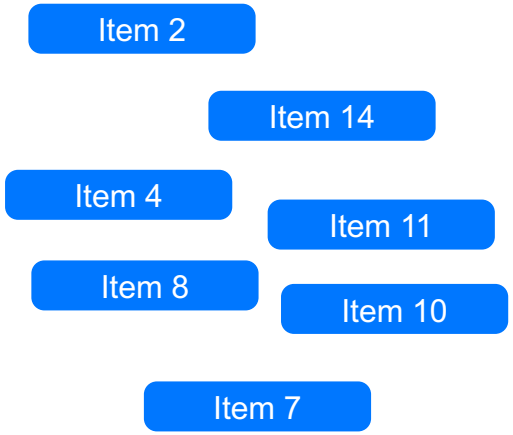


Candidate generation

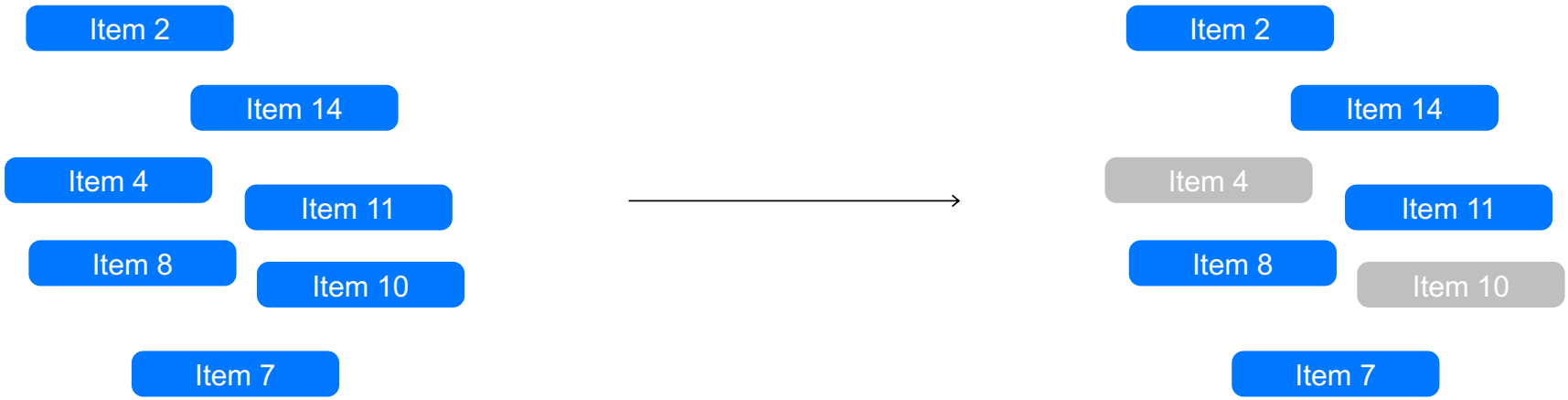


Filtering

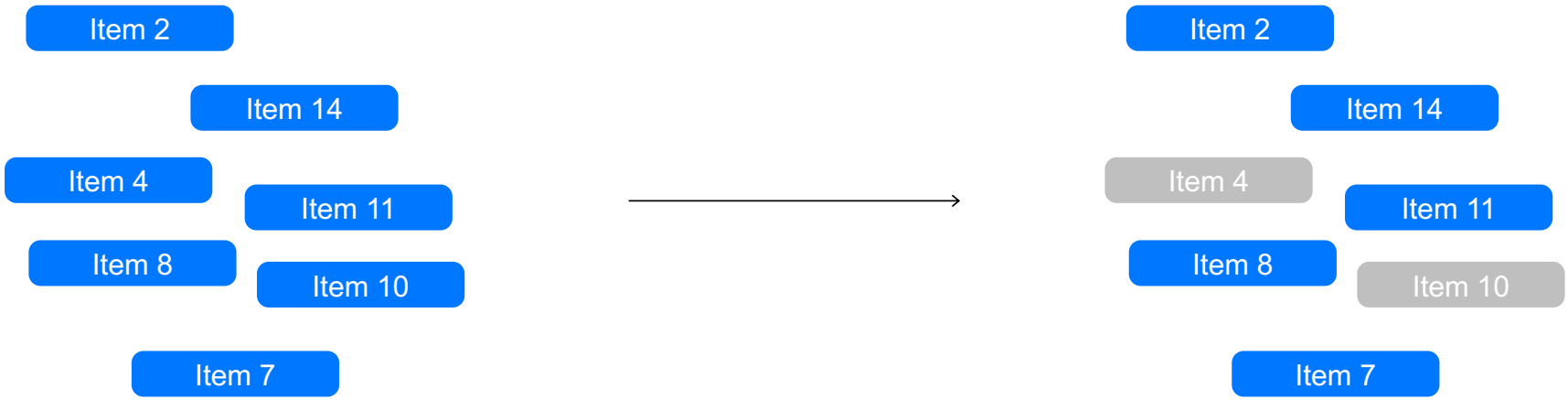
Filtering



Filtering


Item 2
Item 14
Item 4
Item 11
Item 8
Item 10
Item 7
Item 2
Item 14
Item 4
Item 11
Item 8
Item 10
Item 7

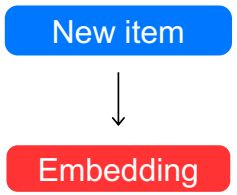
Filtering



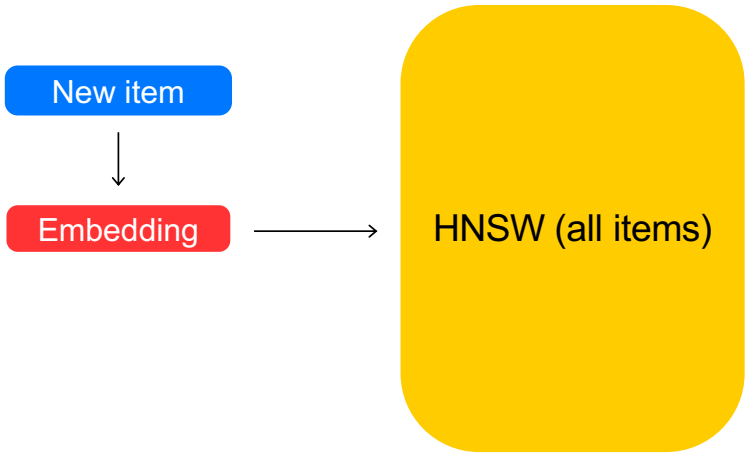
- Static factors
- Light ML models score
- User history

Filtering & candidate generation

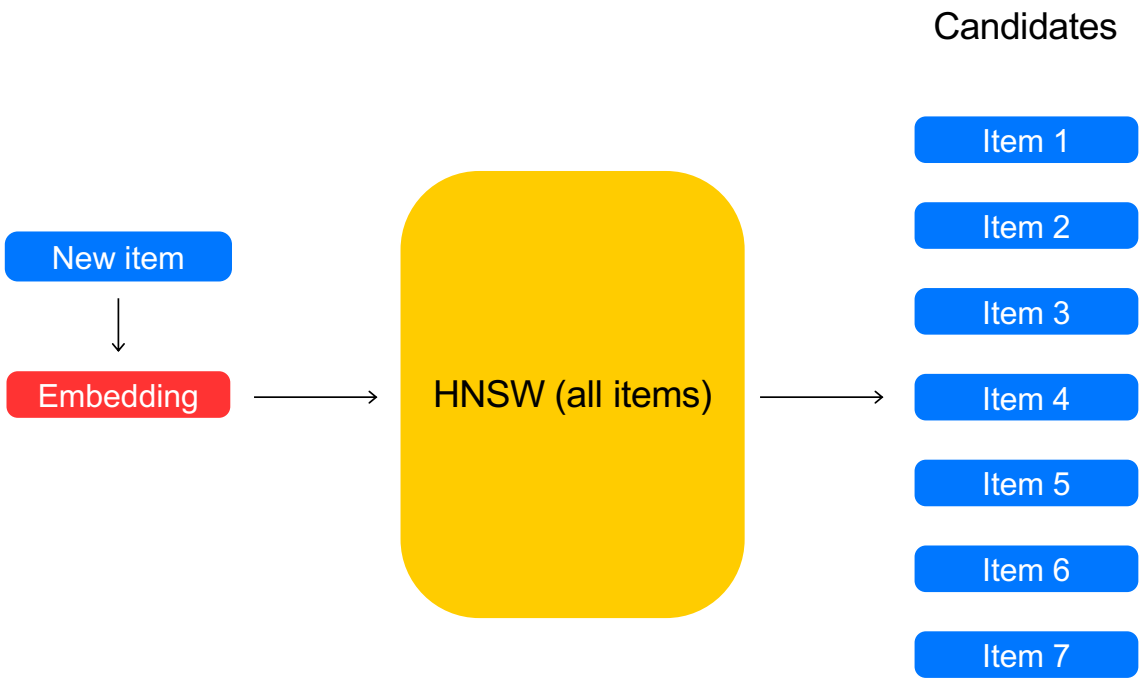
Filtering & candidate generation



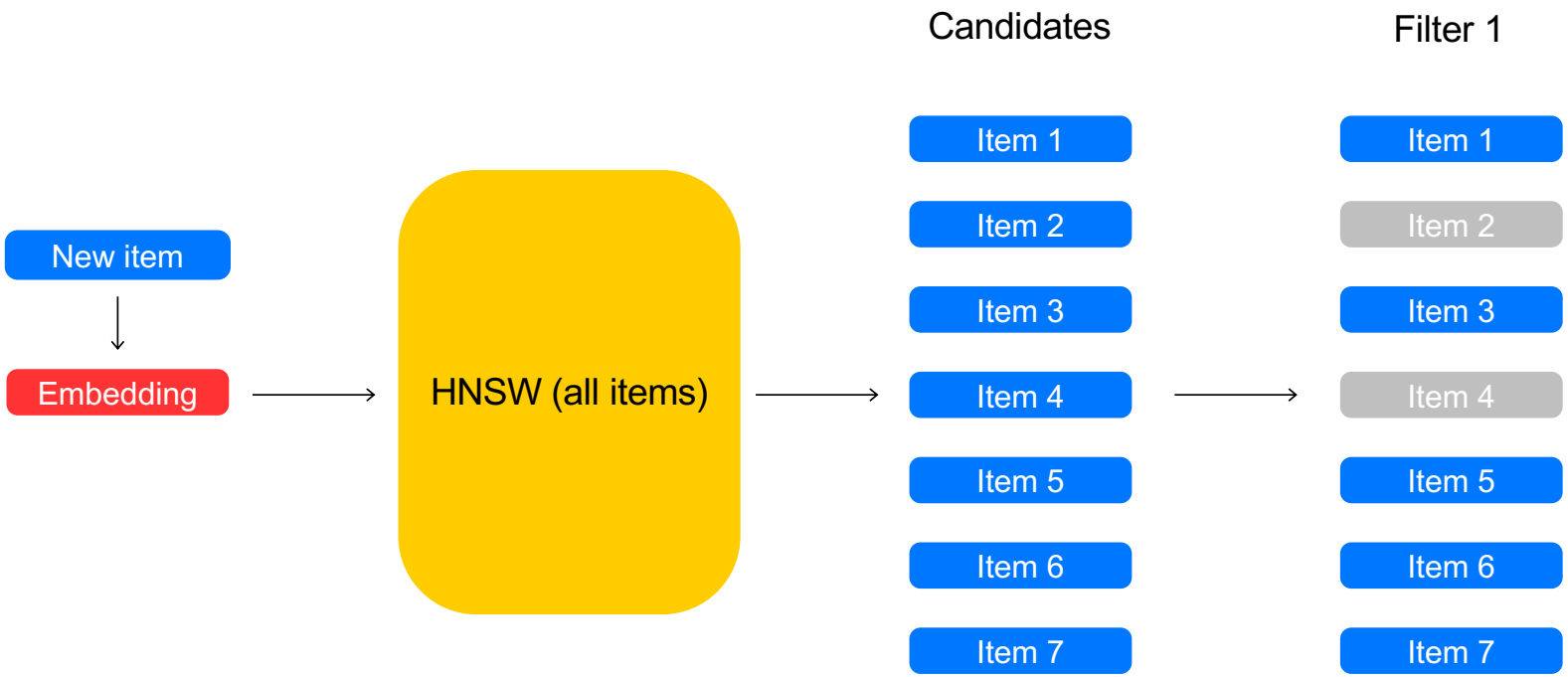
Filtering & candidate generation



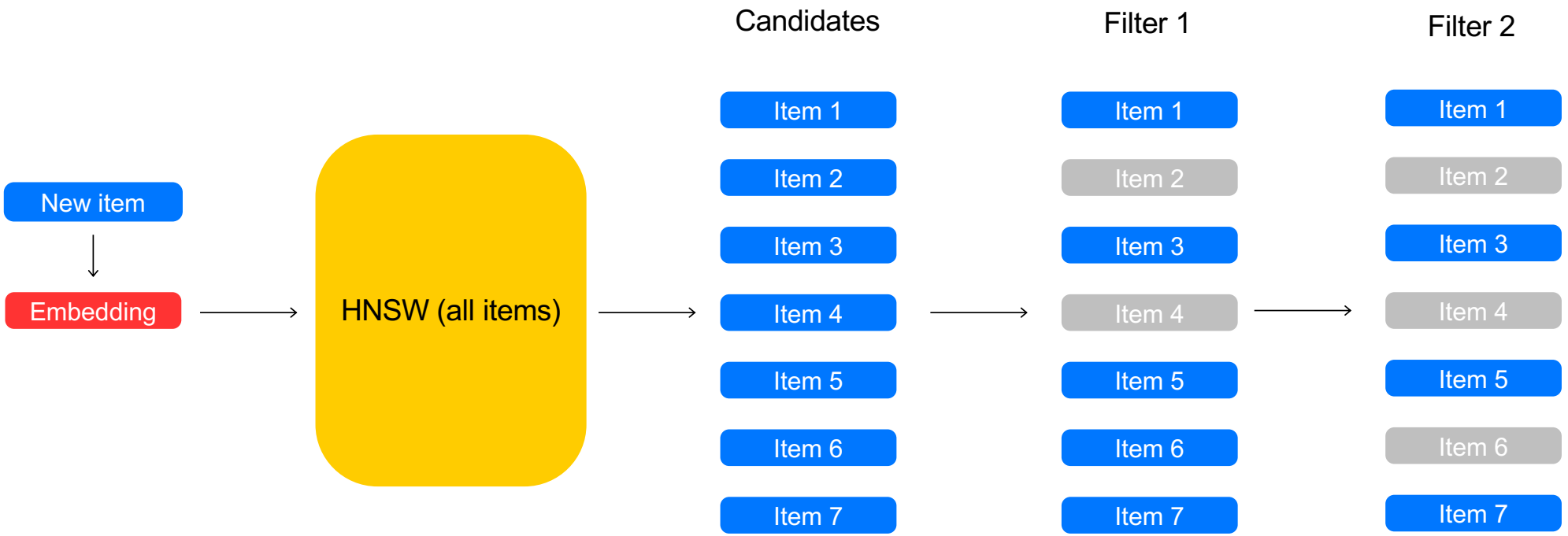
Filtering & candidate generation



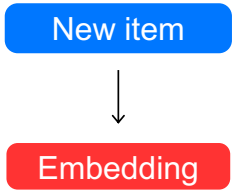
Filtering & candidate generation



Filtering & candidate generation

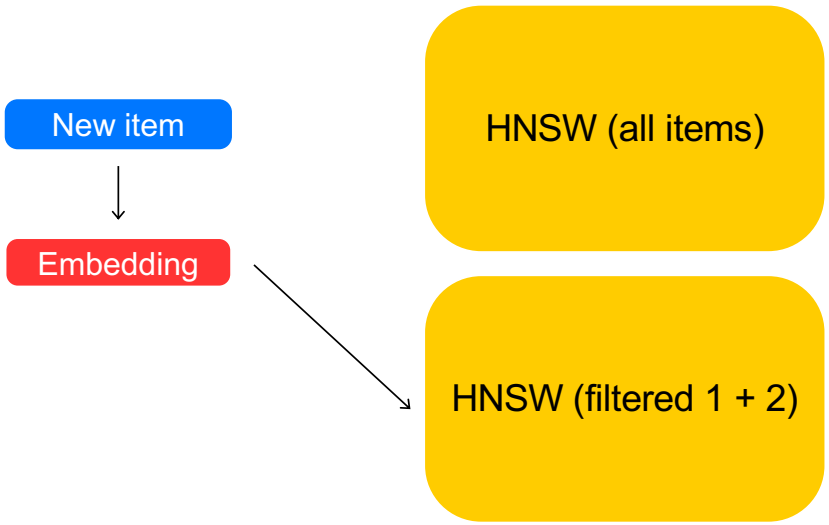


Filtering & candidate generation

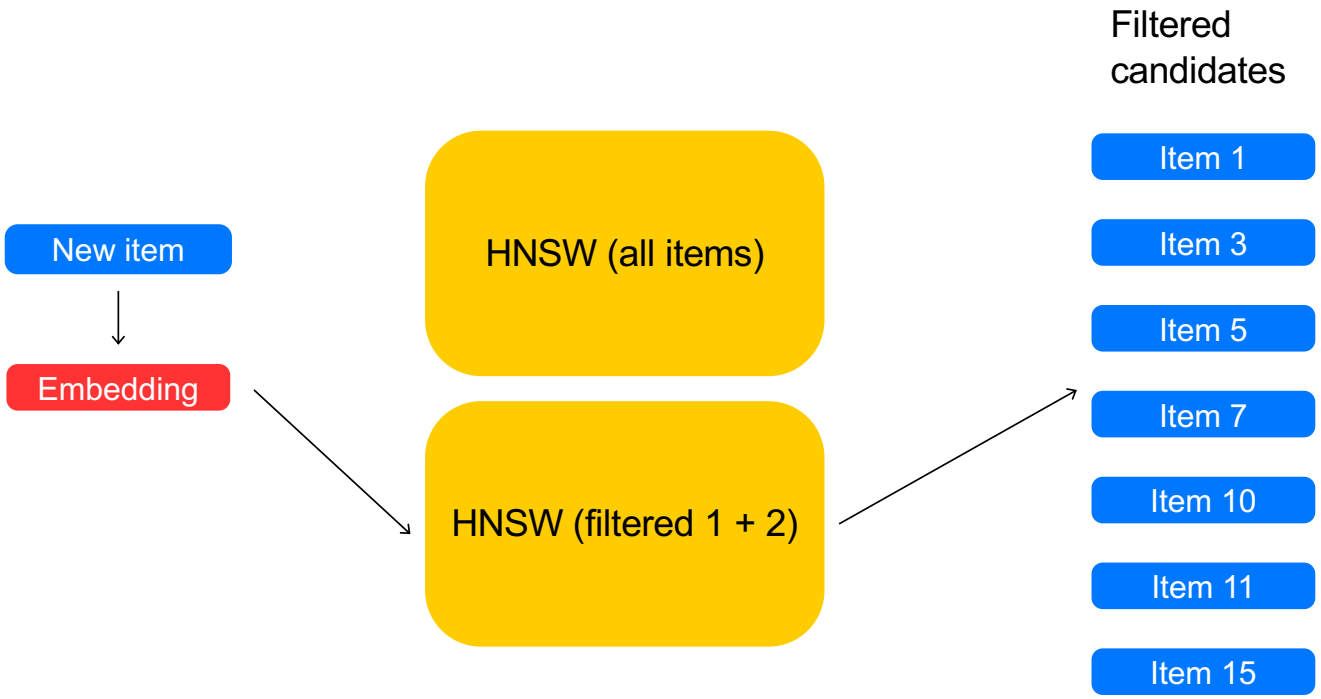


Filtering & candidate generation

Filtering & candidate generation

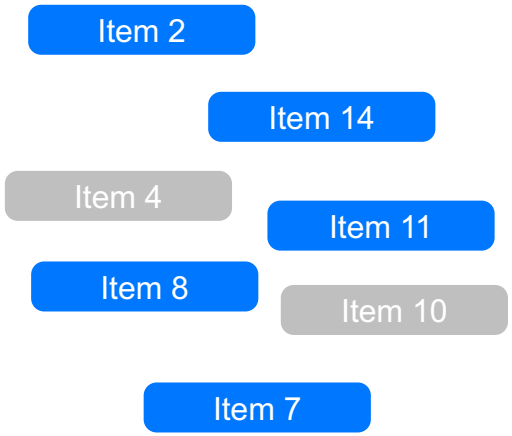


Filtering & candidate generation

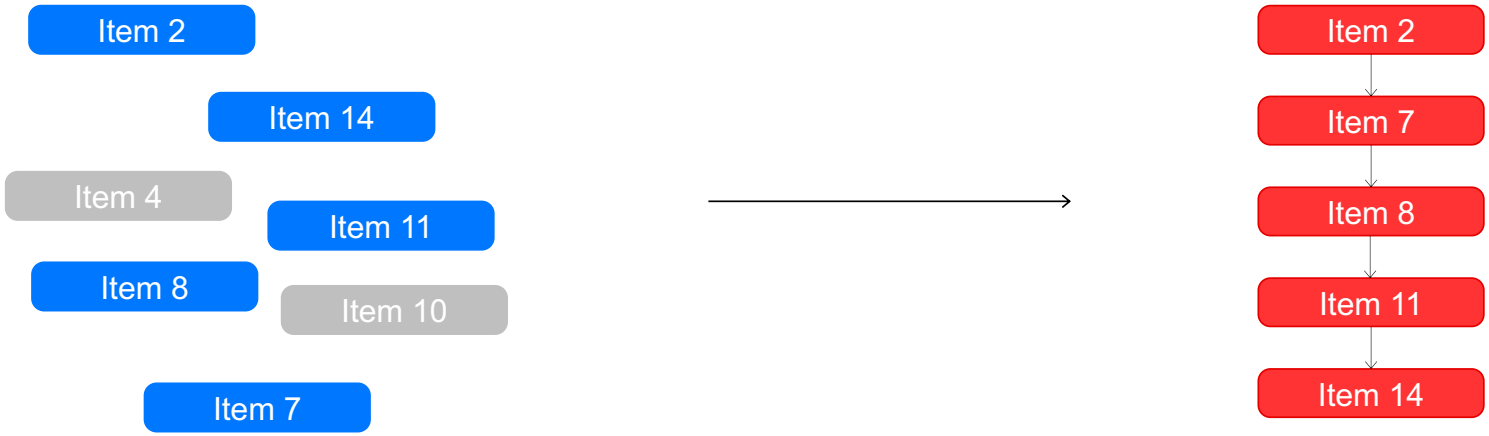


Ranking

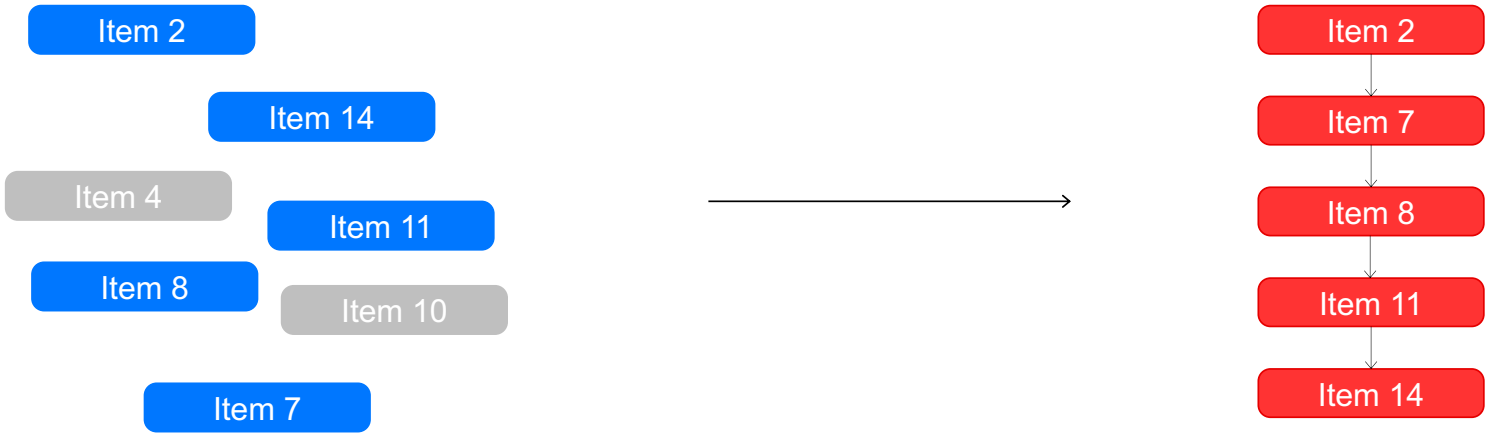
Ranking



Ranking


Item 2
Item 14
Item 4
Item 11
Item 8
Item 10
Item 7
Item 2
Item 7
Item 8
Item 11
Item 14

Ranking

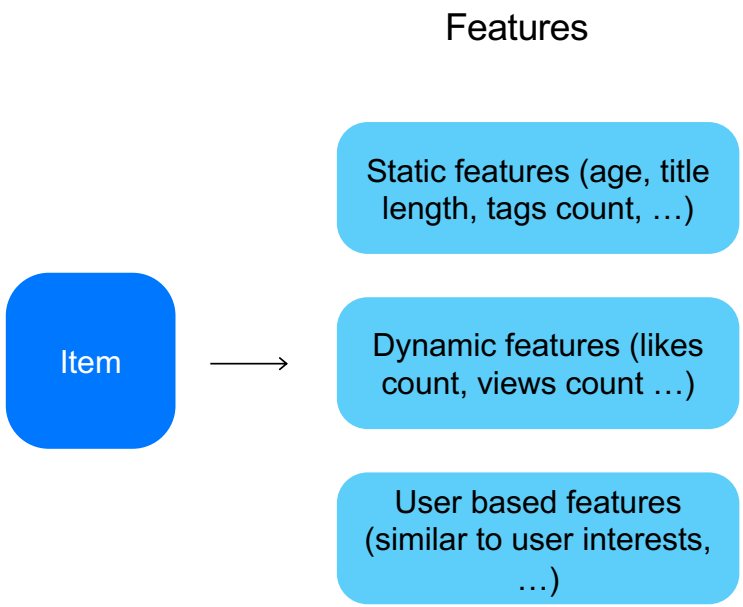


- Static factors
- Heavy ML models score
- Embeddings cosine similarity
- …

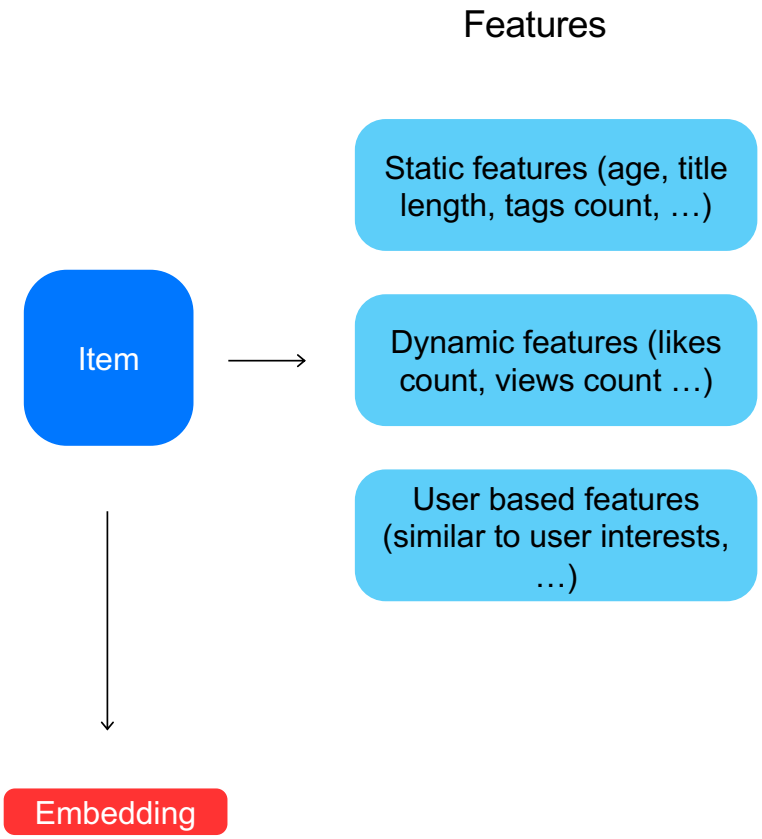
Ranking

Ranking

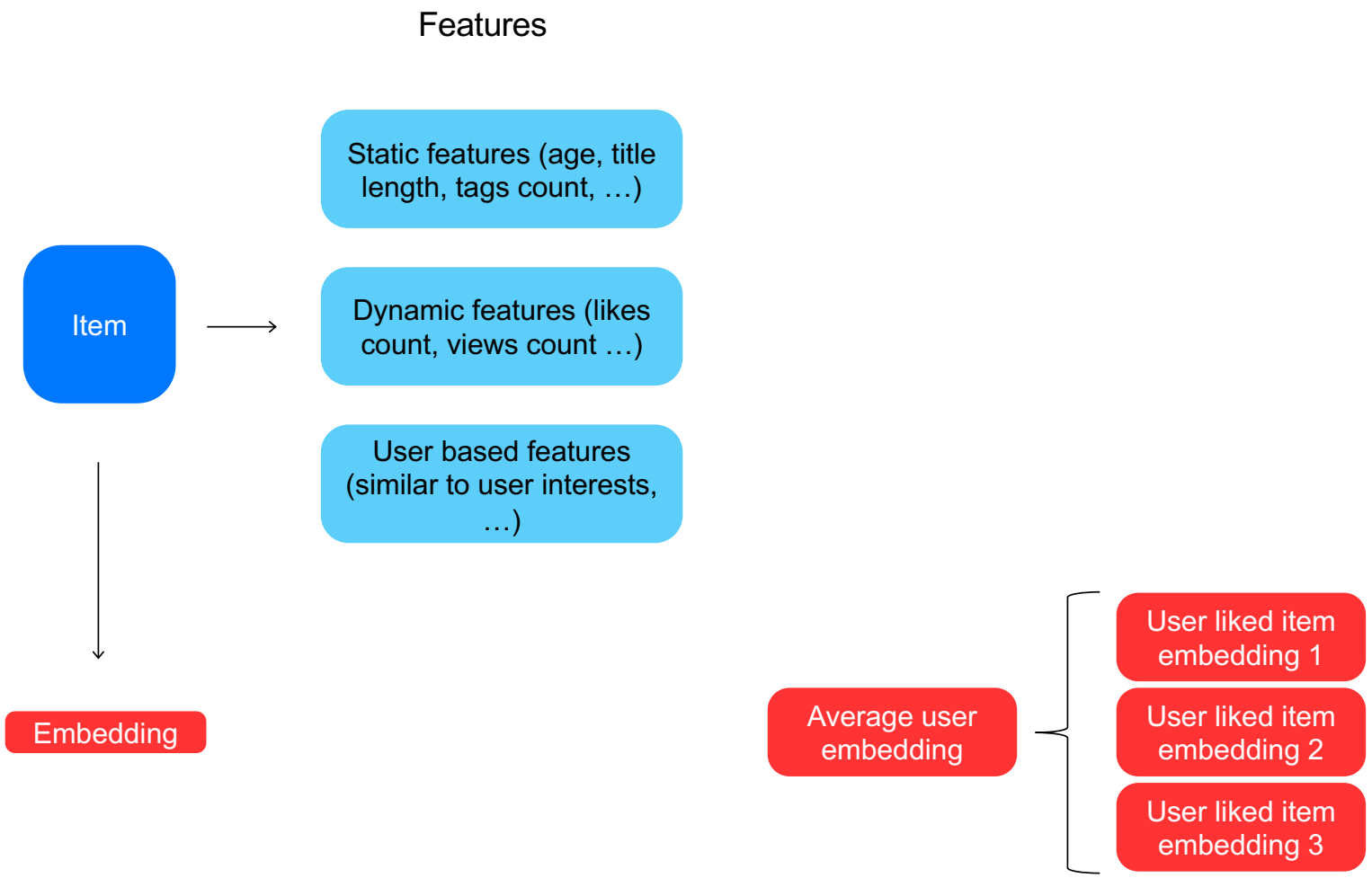
Ranking


Features
Item
Static features (age, title length, tags count, …)
Dynamic features (likes count, views count …)
User based features (similar to user interests, …)

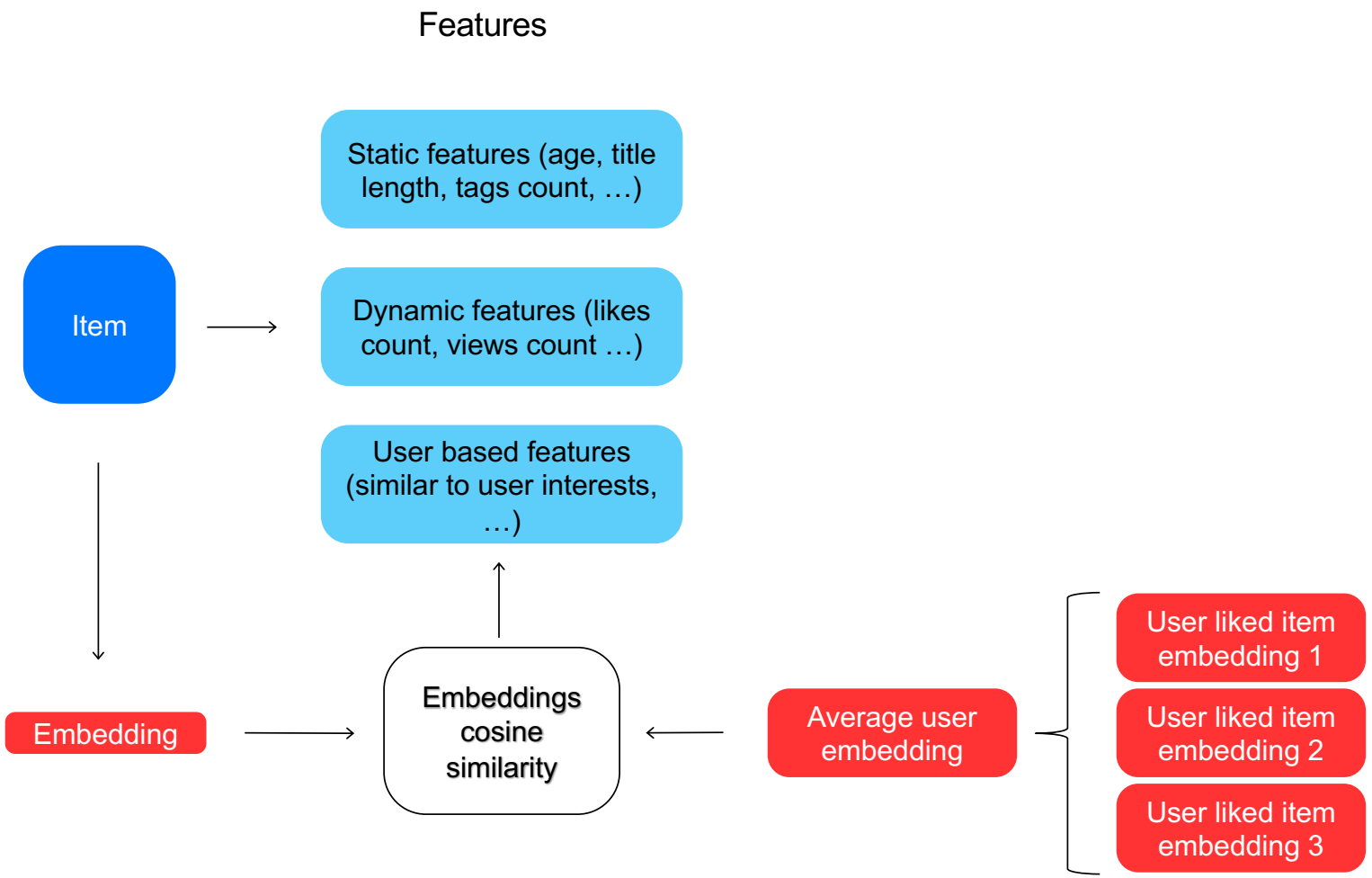
Ranking


Features
Static features (age, title length, tags count, …)
Item
Dynamic features (likes count, views count …)
User based features (similar to user interests, …)
Embedding

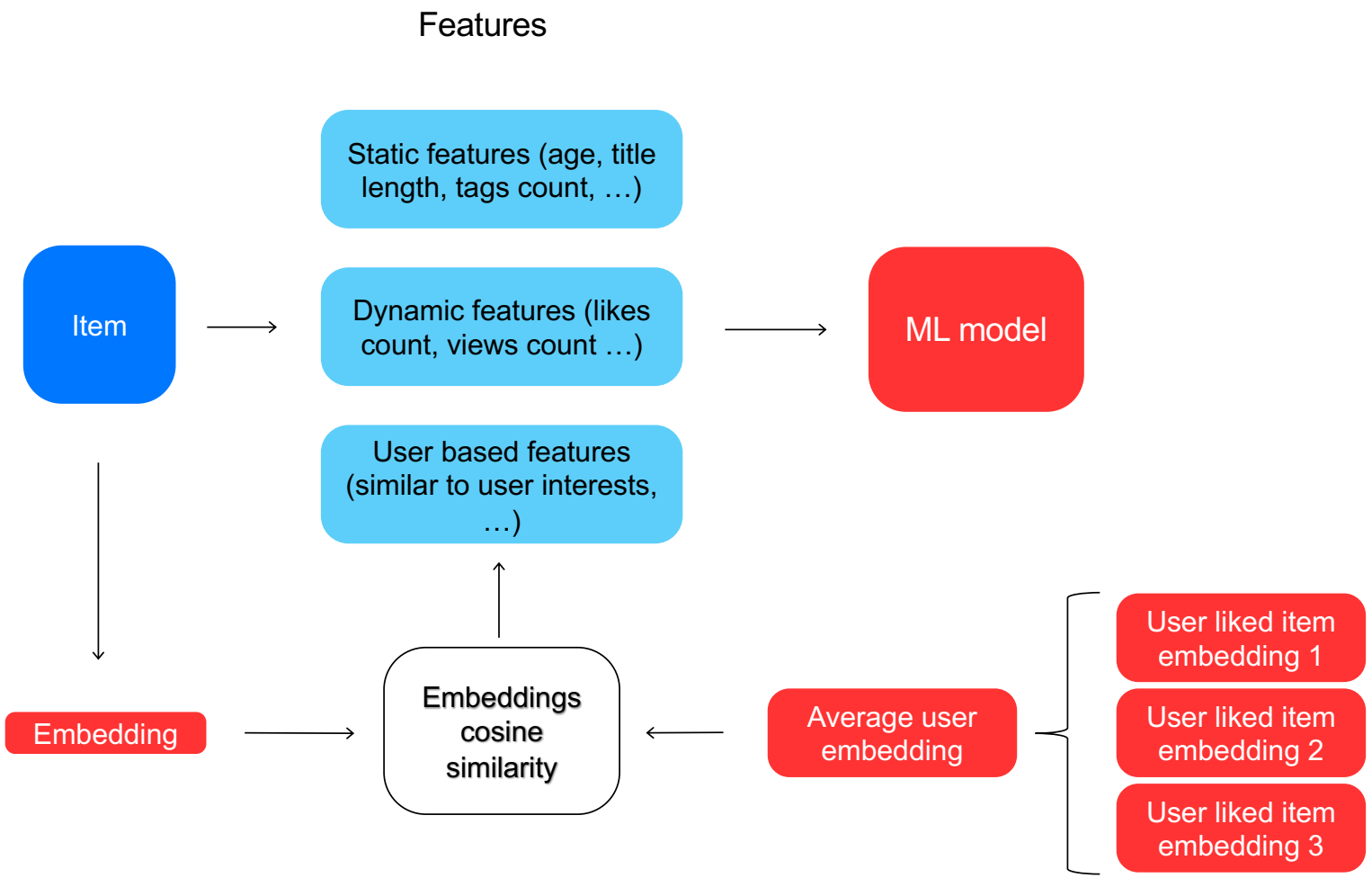
Ranking


Features
Static features (age, title length, tags count, …)
Item
Dynamic features (likes count, views count …)
User based features (similar to user interests, …)
Embedding
Average user embedding
User liked item embedding 1
User liked item embedding 2
User liked item embedding 3

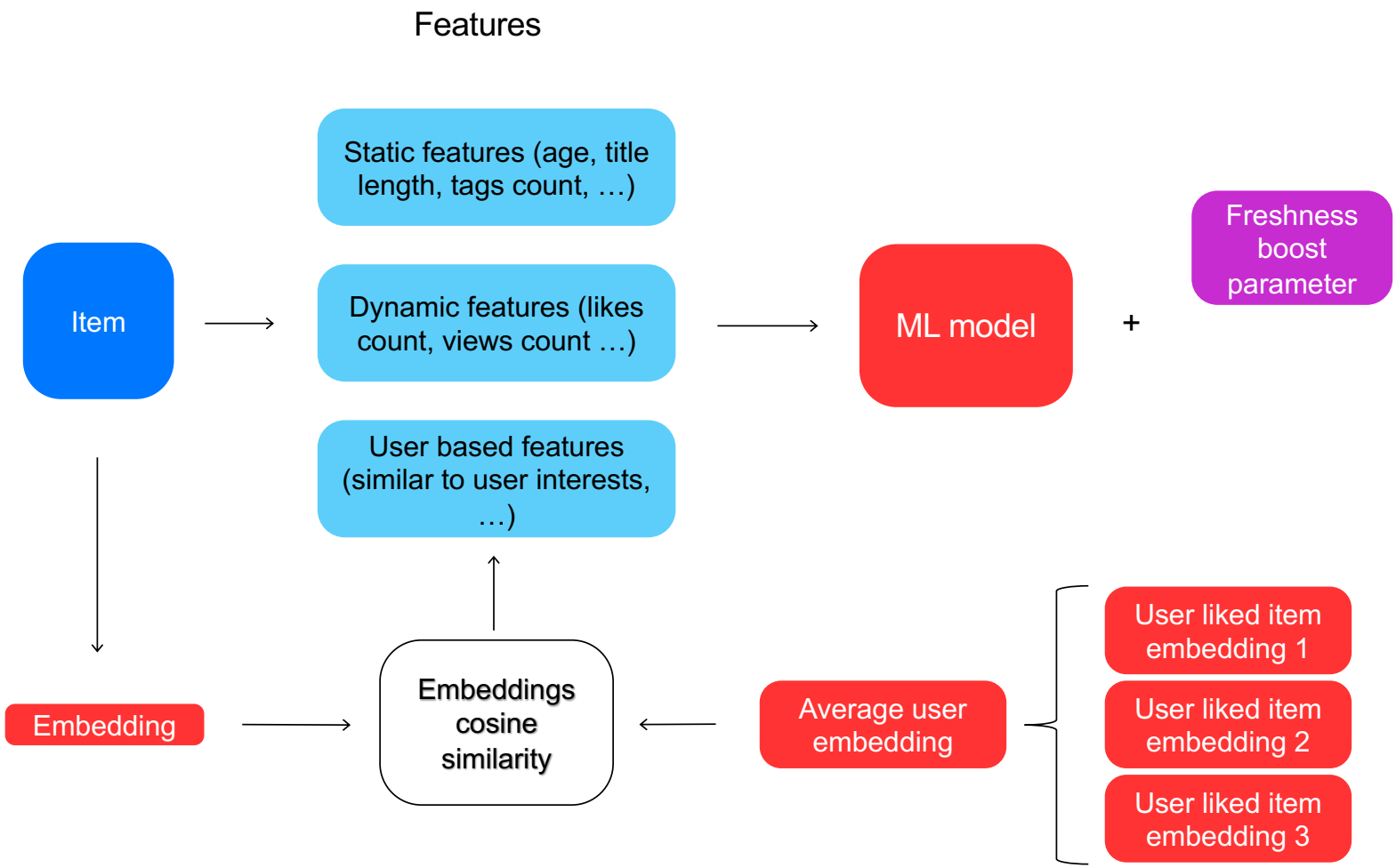
Ranking


Features
Static features (age, title length, tags count, …)
Item
Dynamic features (likes count, views count …)
User based features (similar to user interests, …)
Embedding
Embeddings cosine similarity
Average user embedding
User liked item embedding 1
User liked item embedding 2
User liked item embedding 3

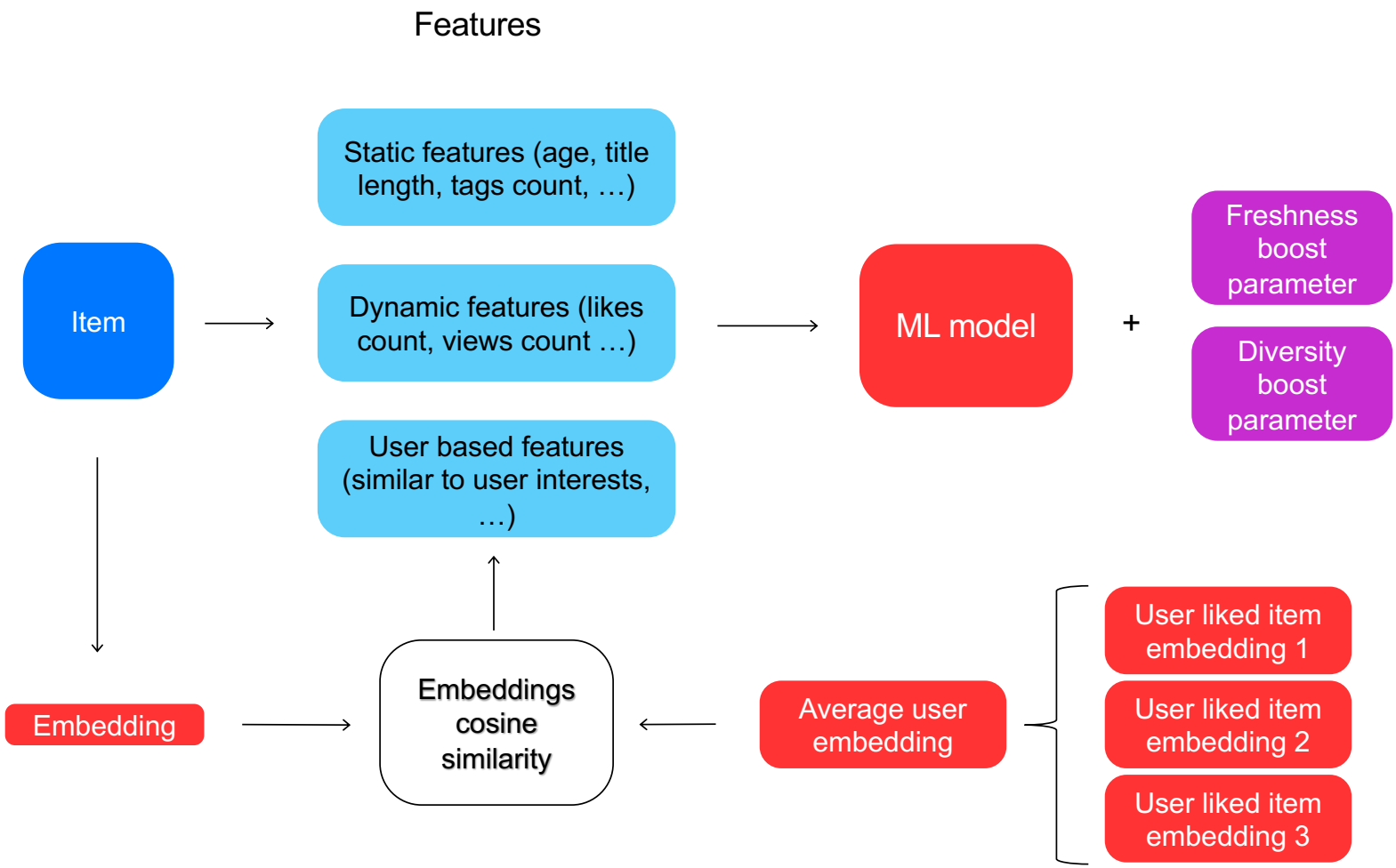
Ranking


Features
Static features (age, title length, tags count, …)
Item
Dynamic features (likes count, views count …)
ML model
User based features (similar to user interests, …)
Embedding
Embeddings cosine similarity
Average user embedding
User liked item embedding 1
User liked item embedding 2
User liked item embedding 3

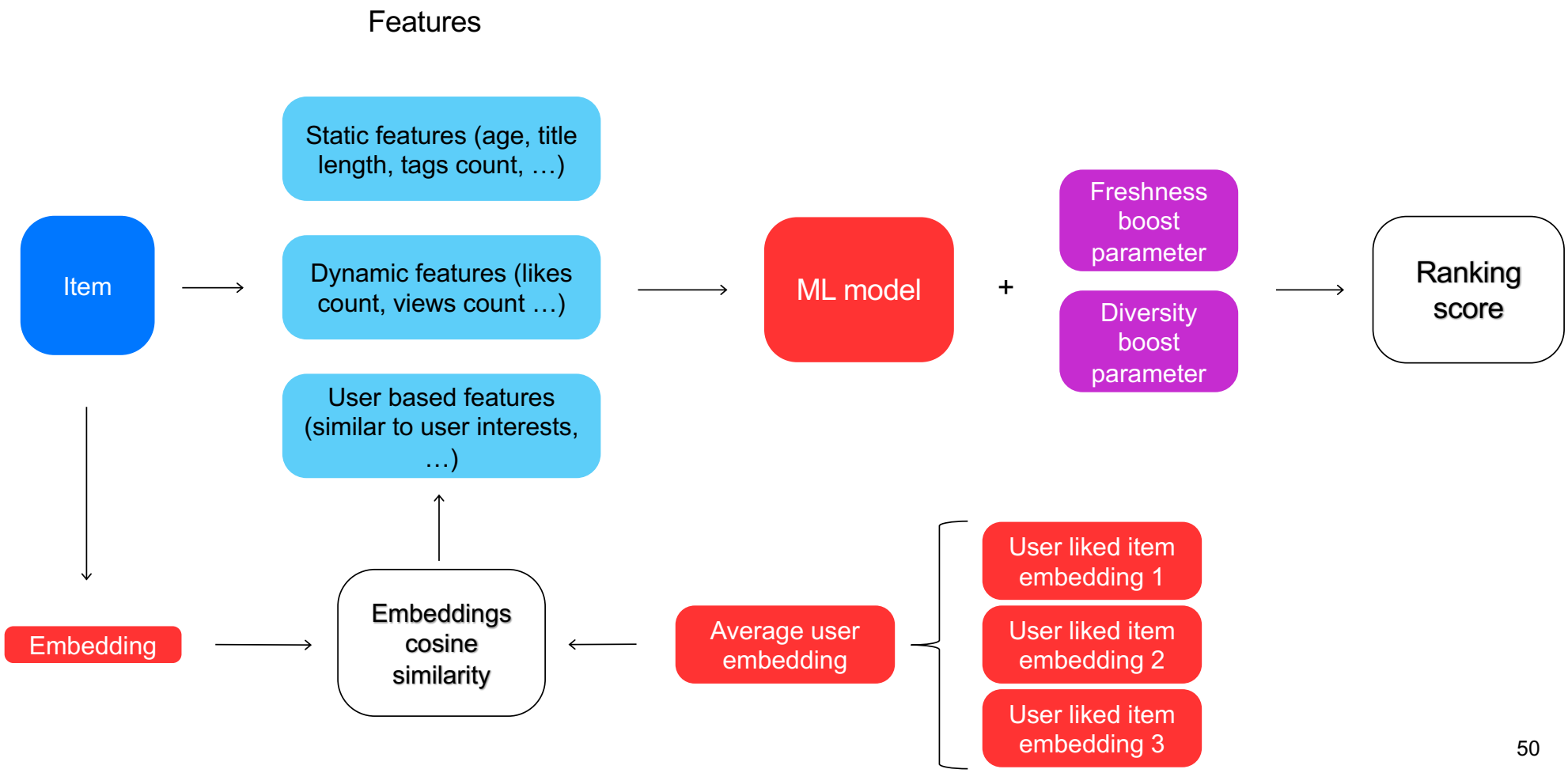
Ranking


Features
Static features (age, title length, tags count, …)
Item
Dynamic features (likes count, views count …)
ML model
+
Freshness boost parameter
User based features (similar to user interests, …)
Embedding
Embeddings cosine similarity
Average user embedding
User liked item embedding 1
User liked item embedding 2
User liked item embedding 3

Ranking


Features
Static features (age, title length, tags count, …)
Item
Dynamic features (likes count, views count …)
ML model
+
Freshness boost parameter
Diversity boost parameter
User based features (similar to user interests, …)
Embedding
Embeddings cosine similarity
Average user embedding
User liked item embedding 1
User liked item embedding 2
User liked item embedding 3

Ranking


Features
Static features (age, title length, tags count, …)
Item
Dynamic features (likes count, views count …)
ML model
+
Freshness boost parameter
Diversity boost parameter
Ranking score
User based features (similar to user interests, …)
Embedding
Embeddings cosine similarity
Average user embedding
User liked item embedding 1
User liked item embedding 2
User liked item embedding 3

02

Examples

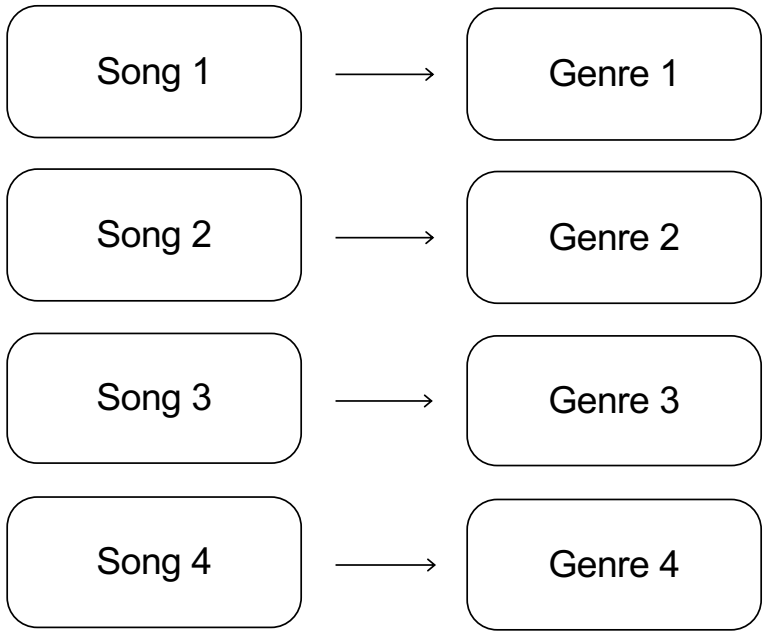
Rubricating

Rubricating

Spotify, Apple Music, Yandex Music, …

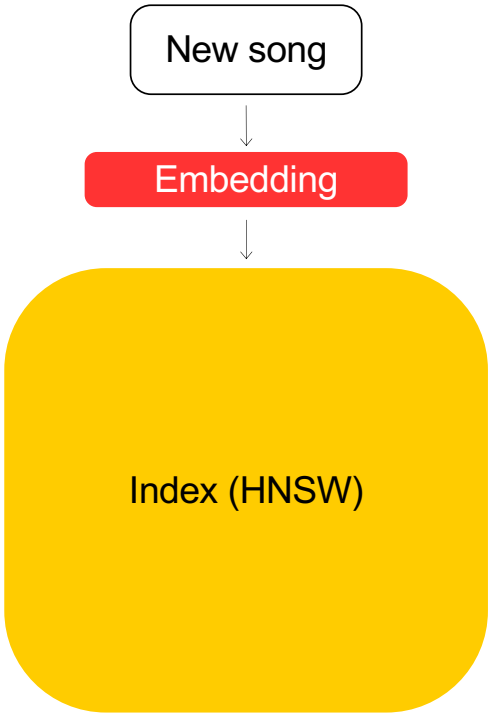
Rubricating

Spotify, Apple Music, Yandex Music, …

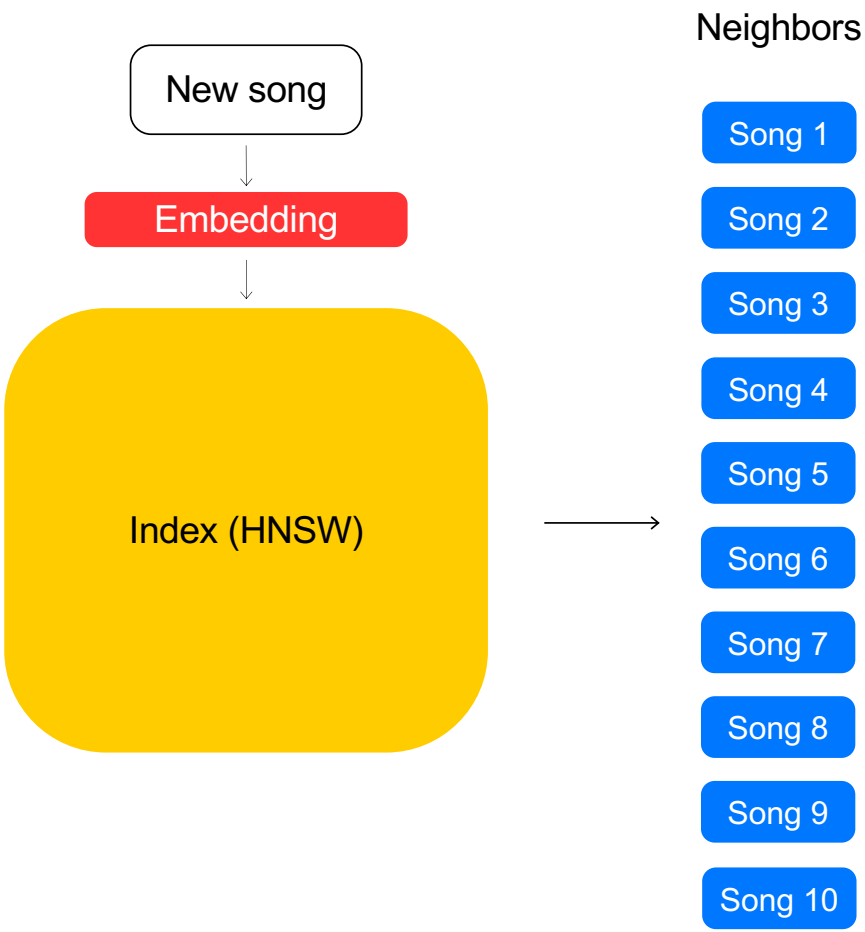


Rubricating

Rubricating

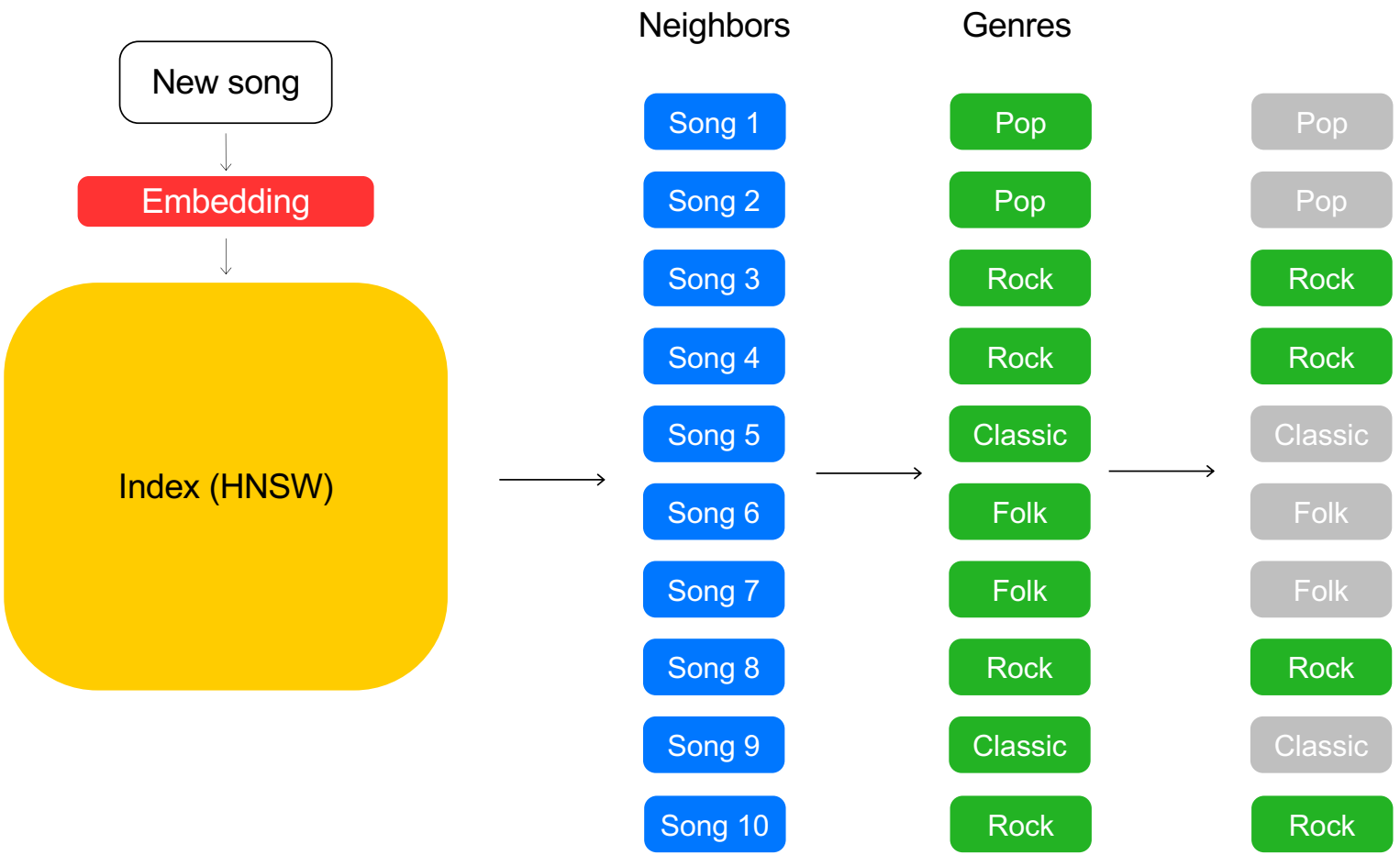


Rubricating

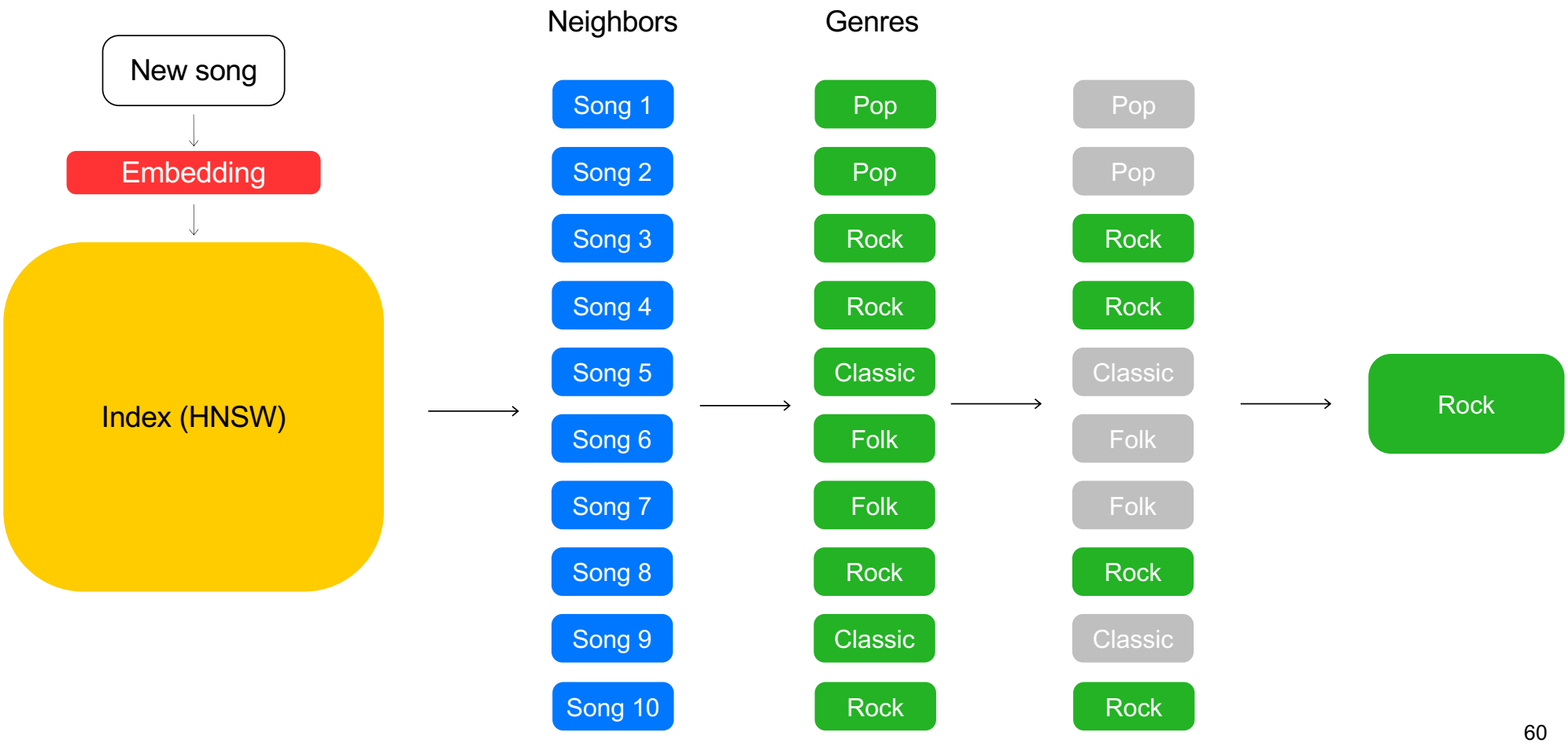


Rubricating

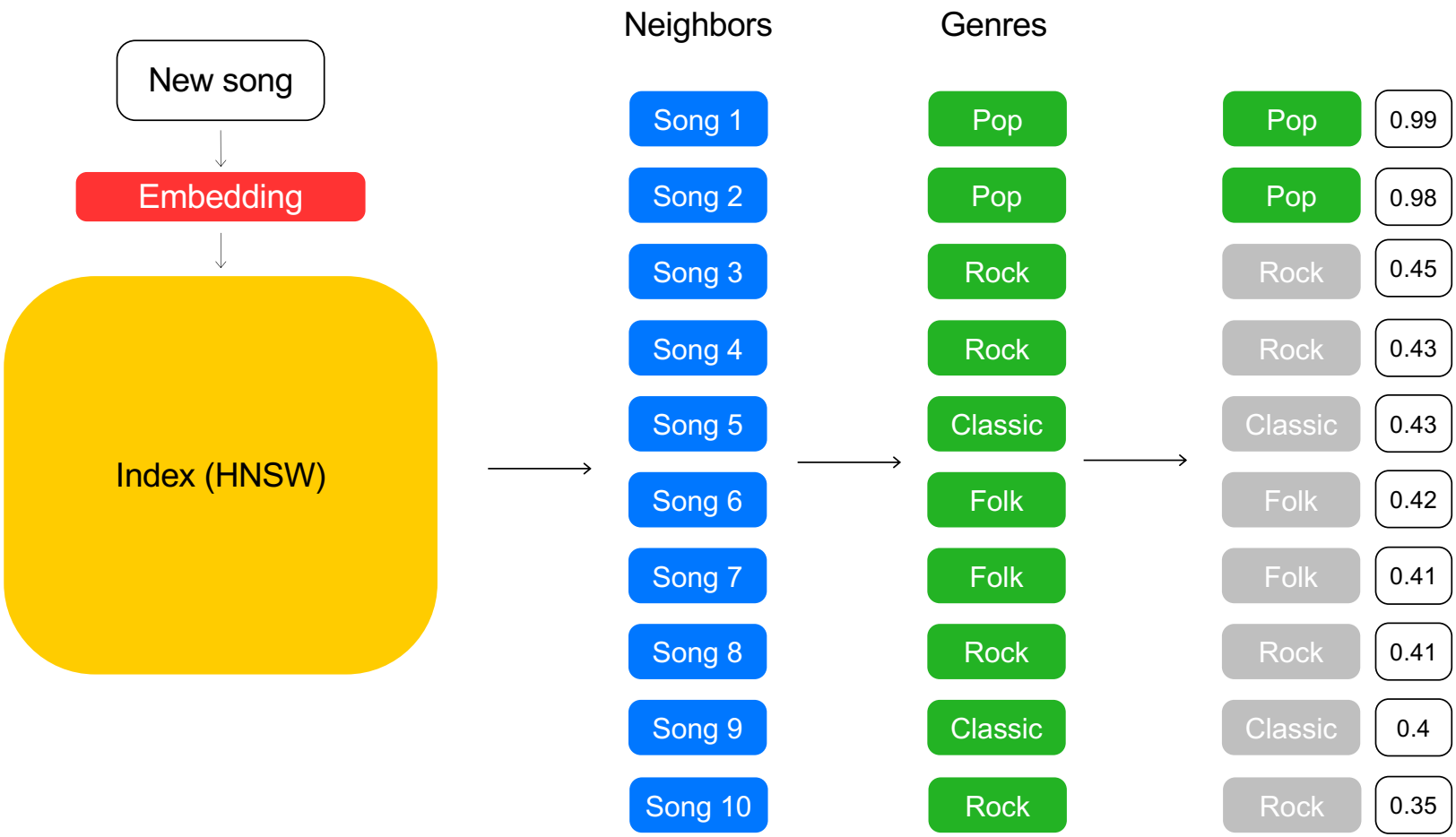
Rubricating



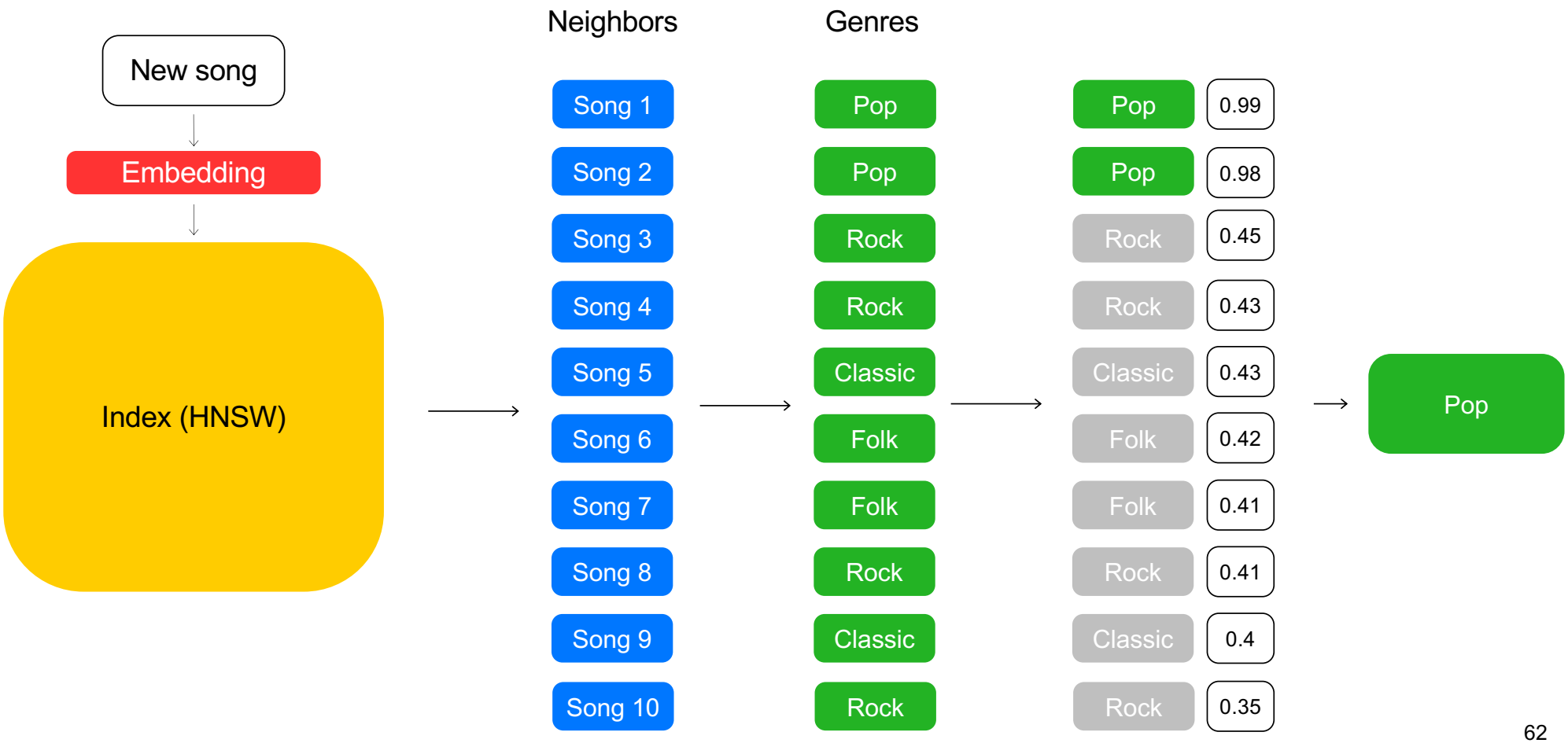
Rubricating



Rubricating



Rubricating



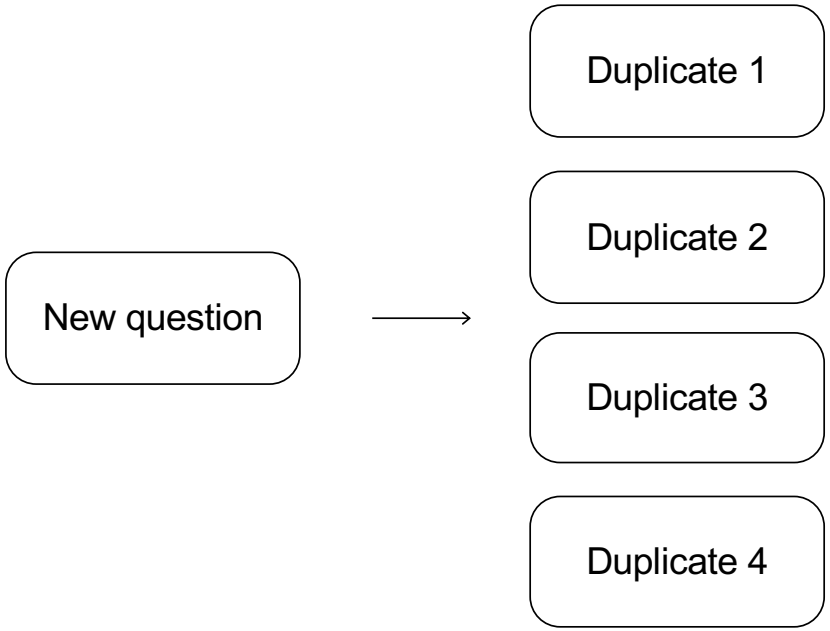
Deduplication

Deduplication

Quora, Yandex Q, …

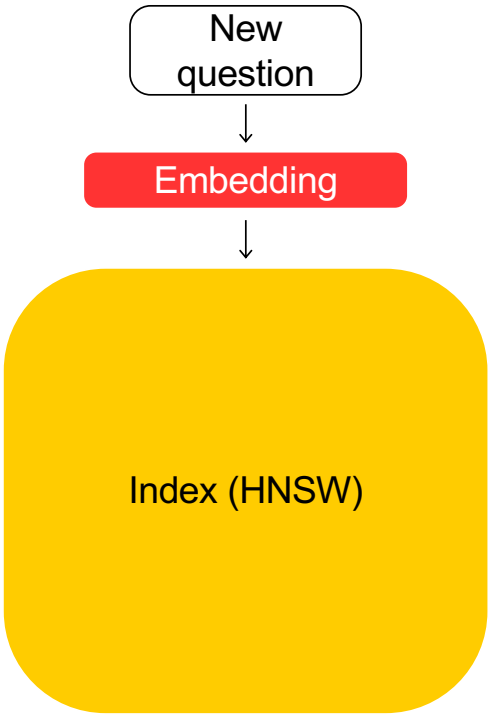
Deduplication

Quora, Yandex.Q, …

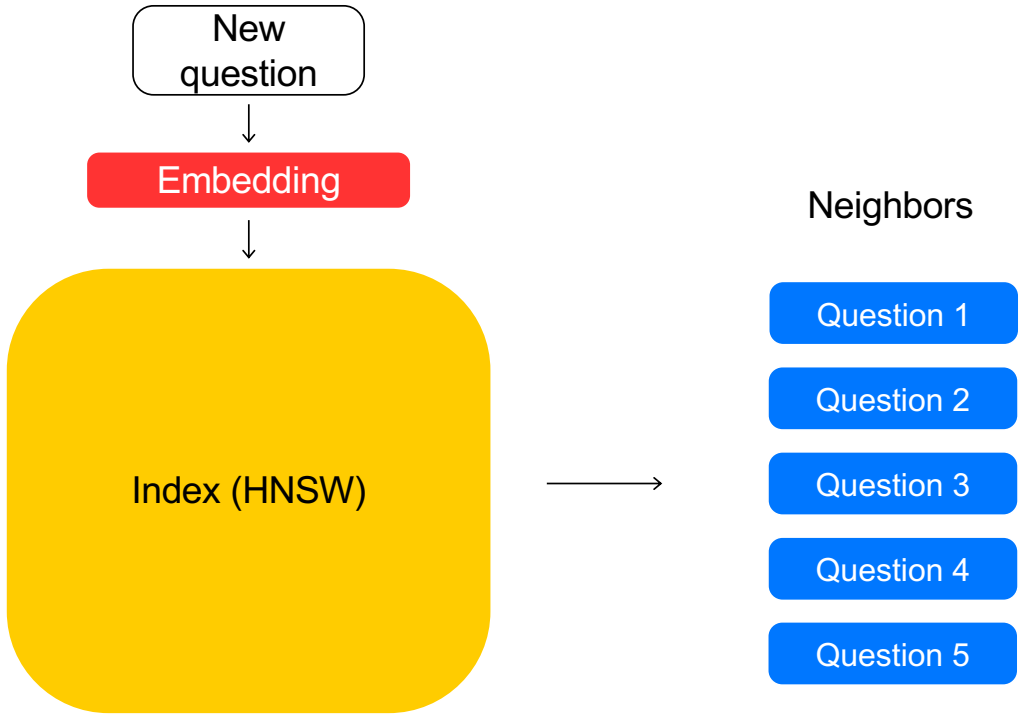


Deduplication

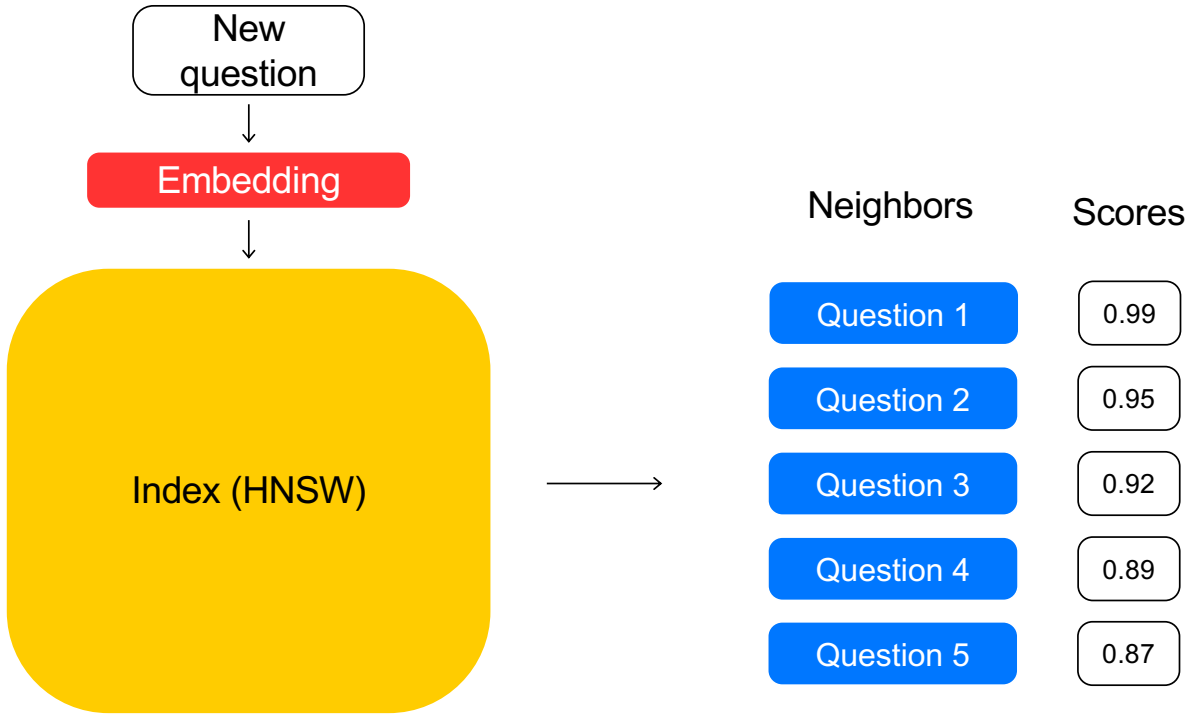
Deduplication



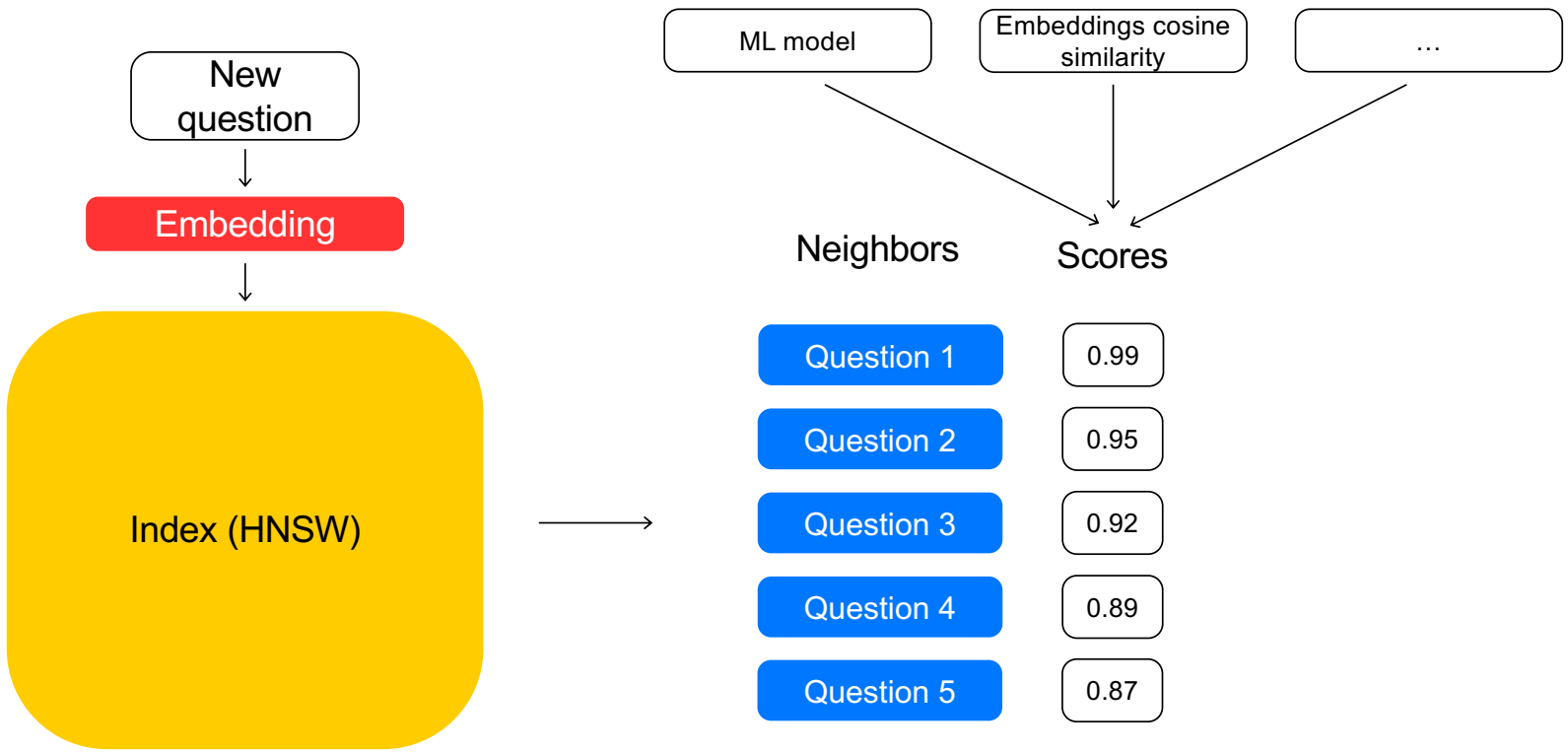
Deduplication



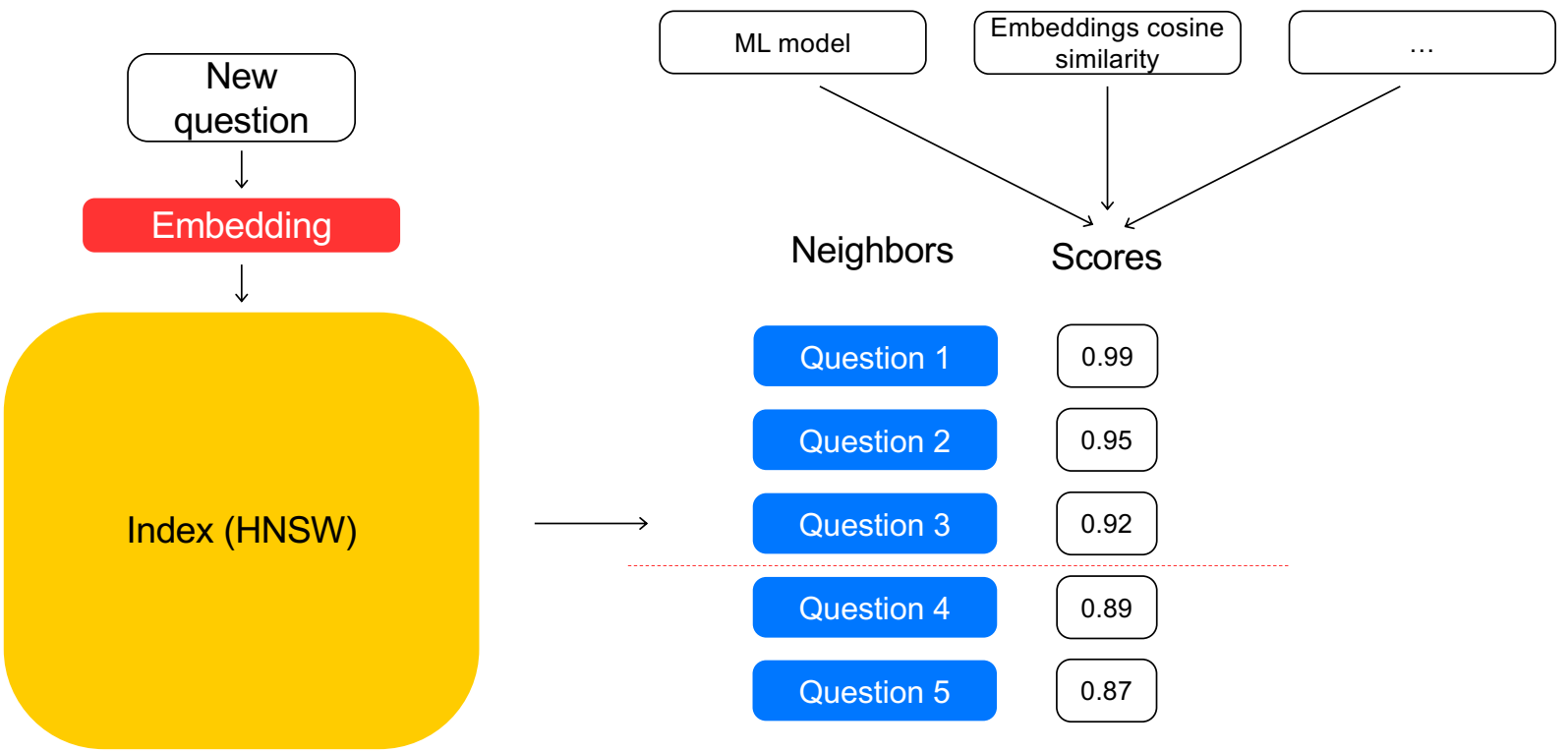
Deduplication



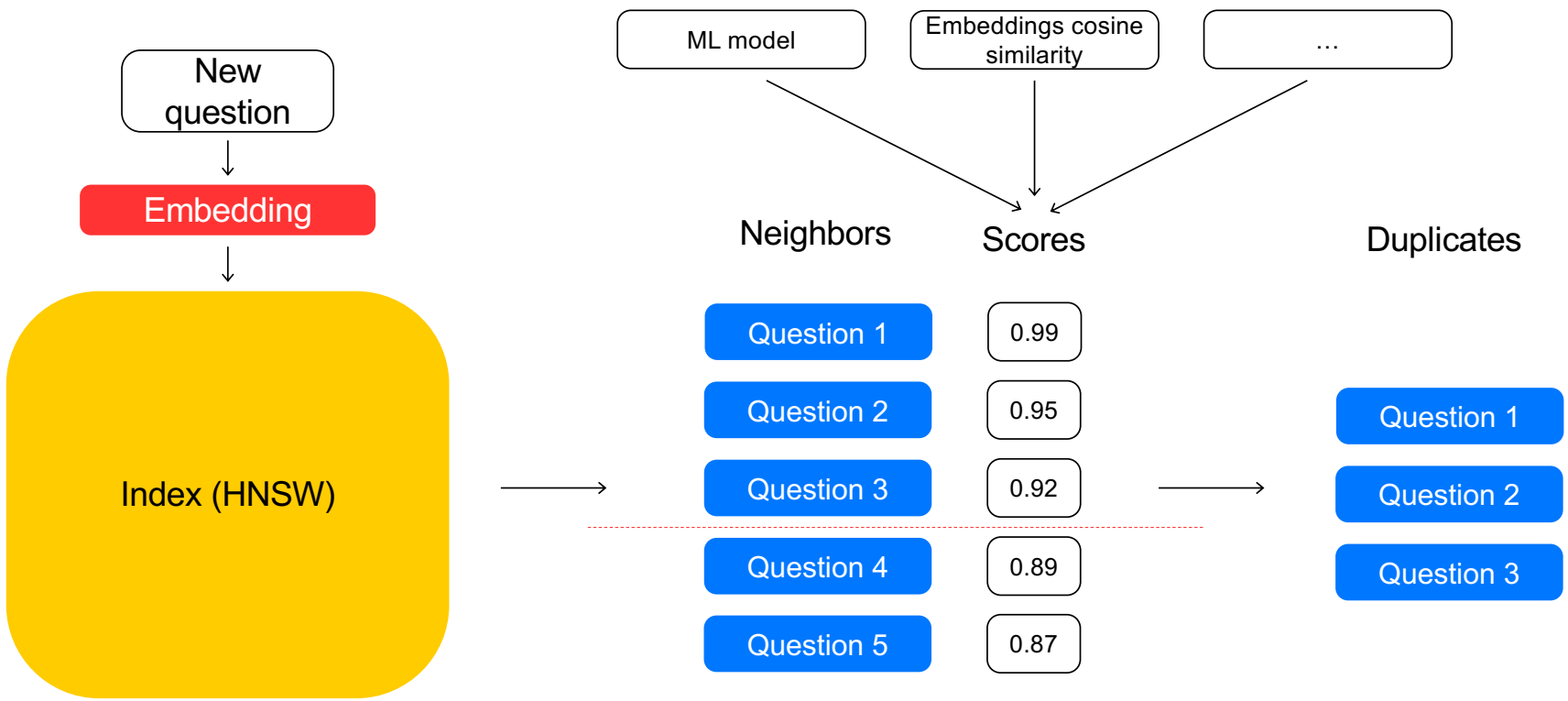
Deduplication



Deduplication



Deduplication



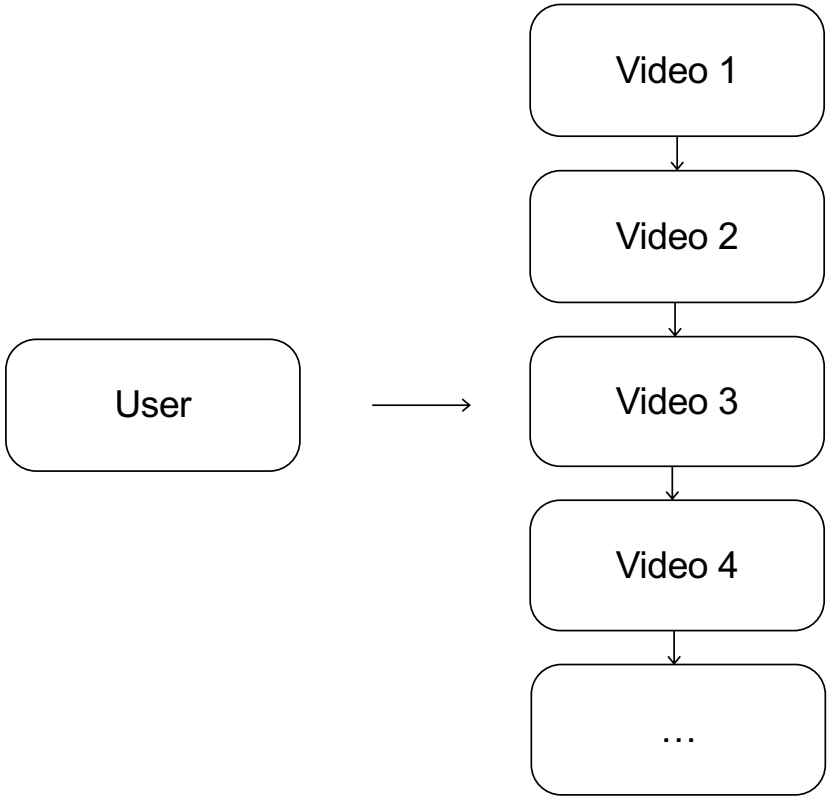
Feed

Feed

YouTube, Netflix, Kinopoisk, …

Feed

YouTube, Netflix, Kinopoisk, …

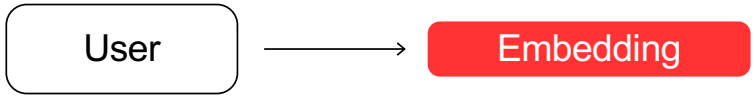


Feed

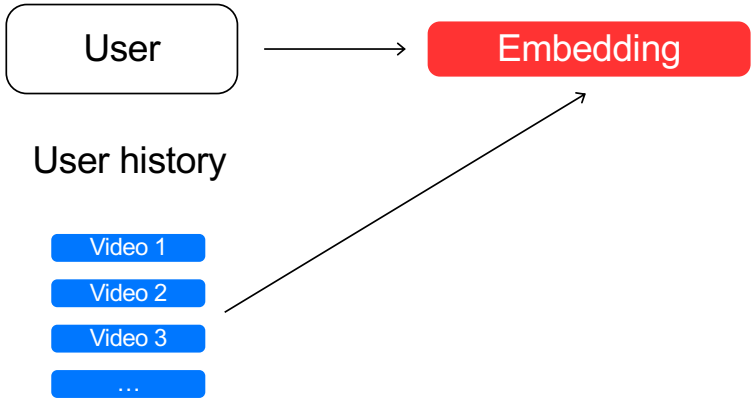
Feed

User

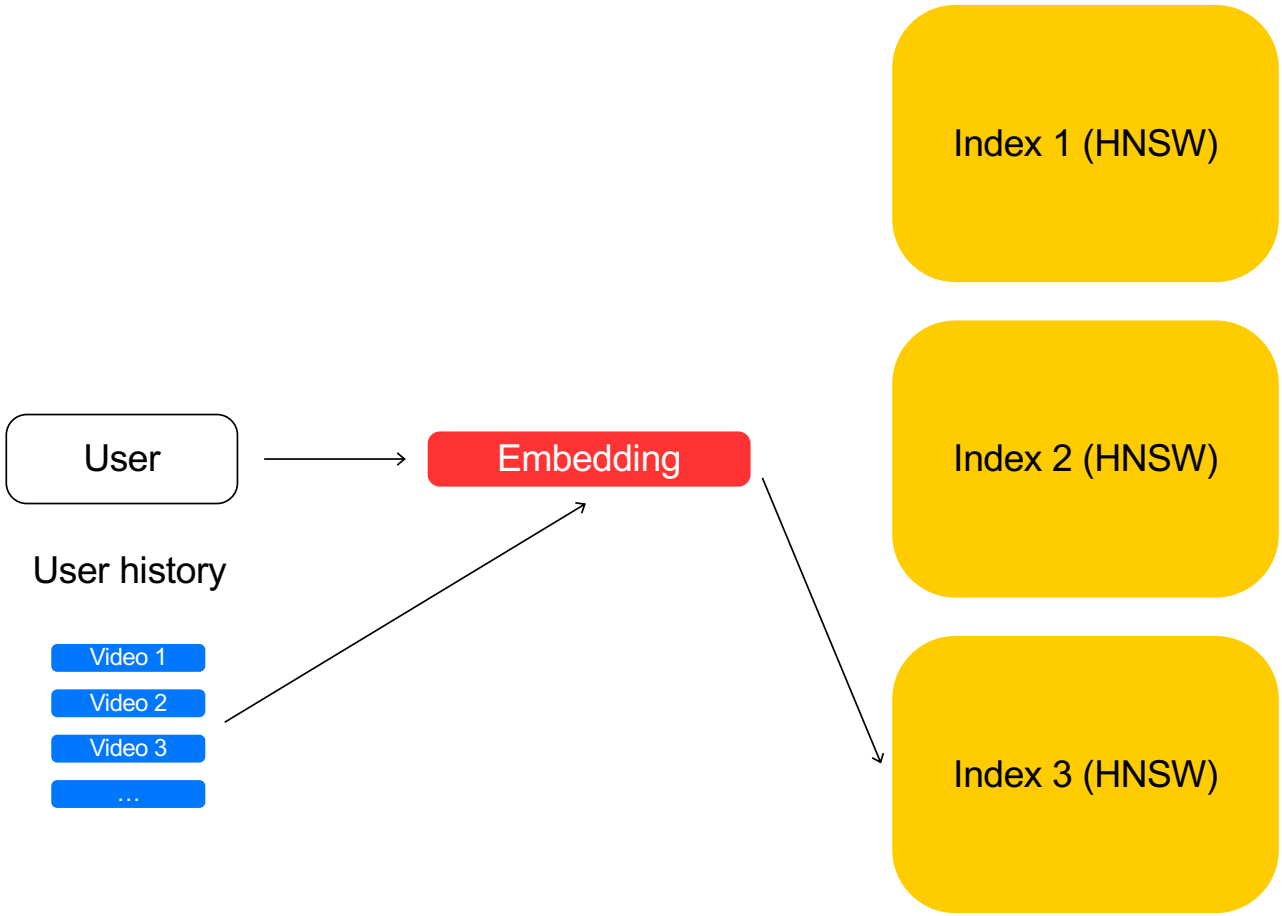
Feed



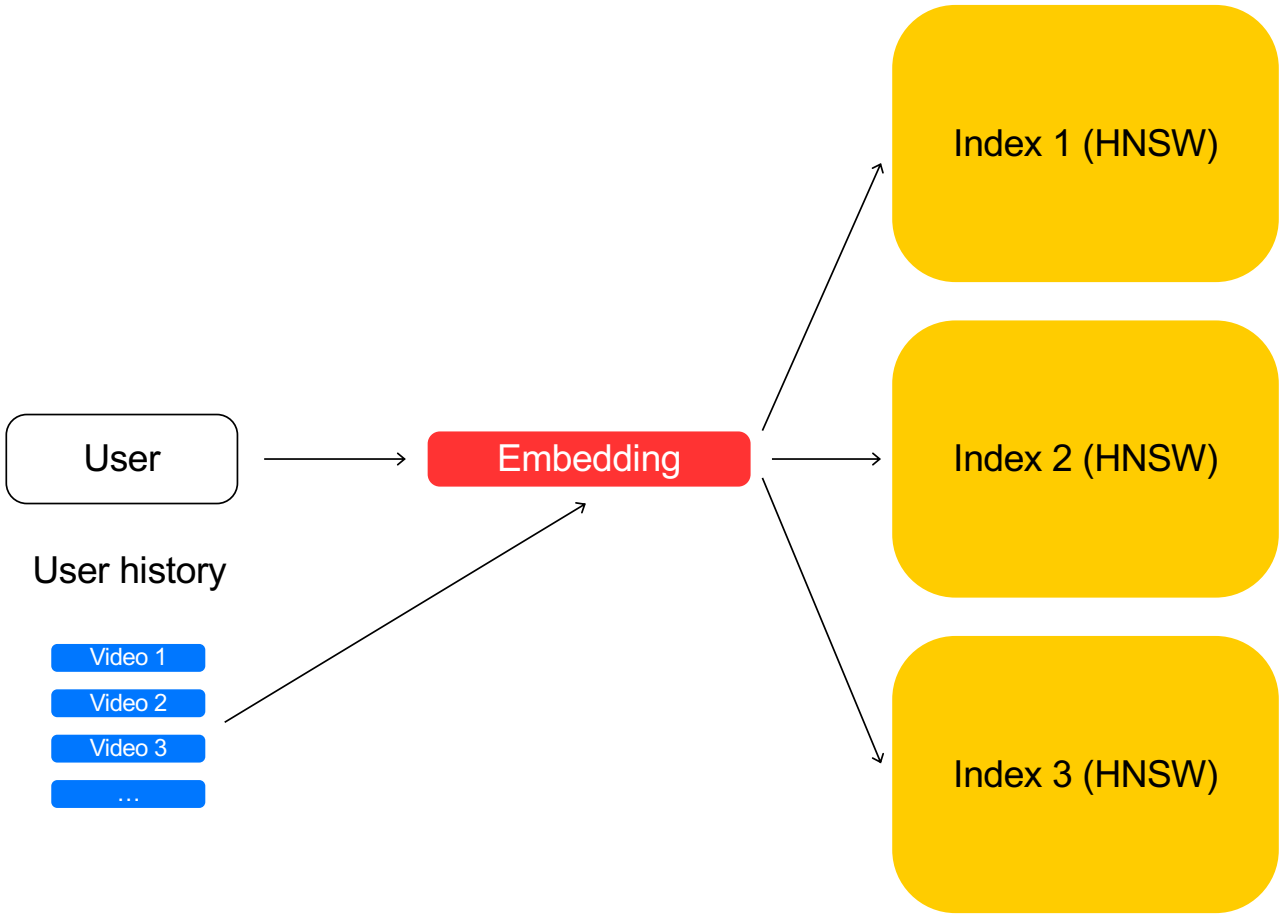
Feed



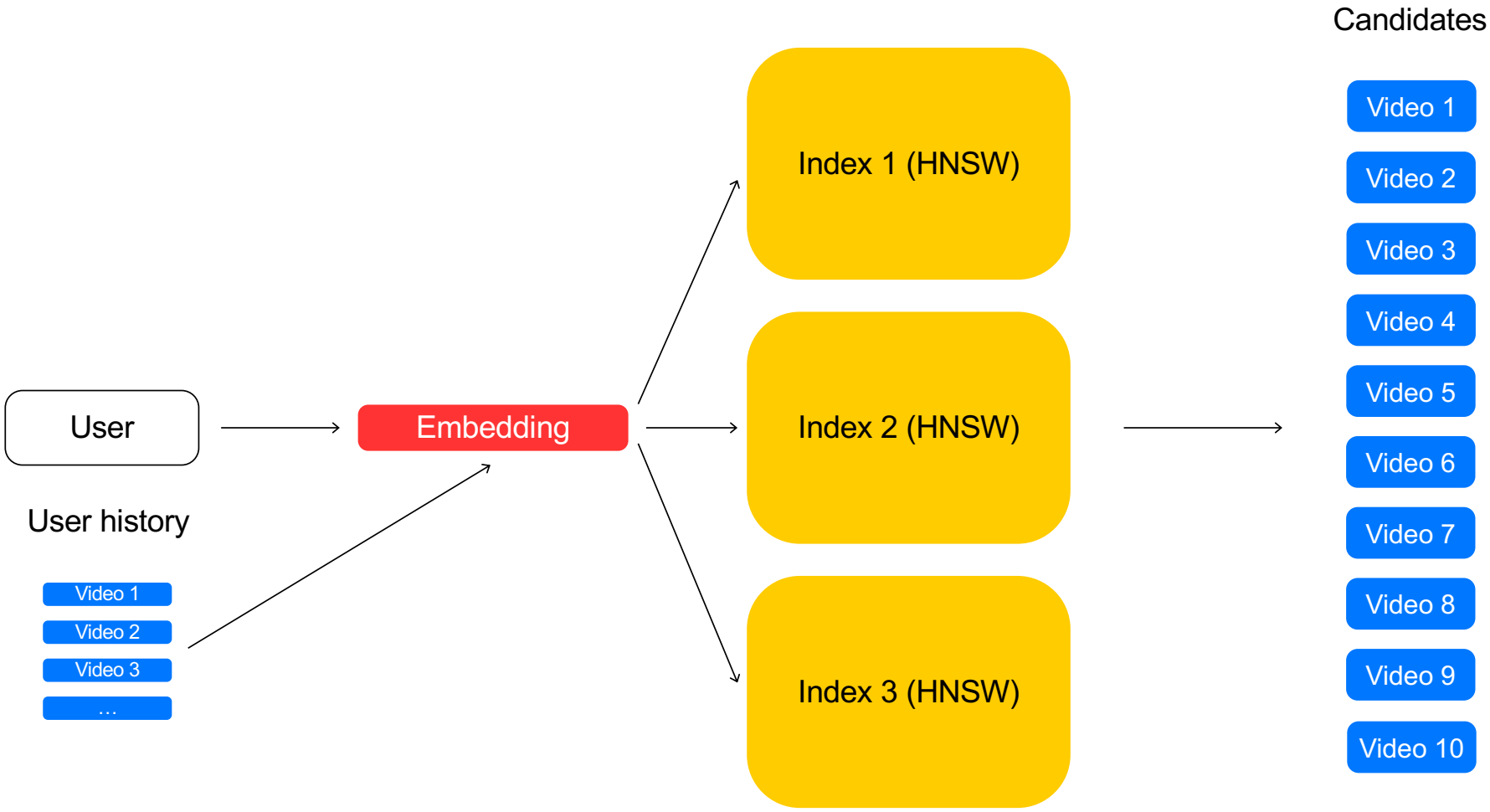
Feed


User
Embedding
User history
Video 1
Video 2
Video 3
...
Index 1 (HNSW)
Index 2 (HNSW)
Index 3 (HNSW)

Feed


User
Embedding
User history
Video 1
Video 2
Video 3
...
Index 1 (HNSW)
Index 2 (HNSW)
Index 3 (HNSW)

Feed


Candidates
User
Embedding
User history
Video 1
Video 2
Video 3
...
Index 1 (HNSW)
Index 2 (HNSW)
Index 3 (HNSW)
Video 1
Video 2
Video 3
Video 4
Video 5
Video 6
Video 7
Video 8
Video 9
Video 10

Feed

Feed

Candidates

Video 1

Video 2

Video 3

Video 4

Video 5

Video 6

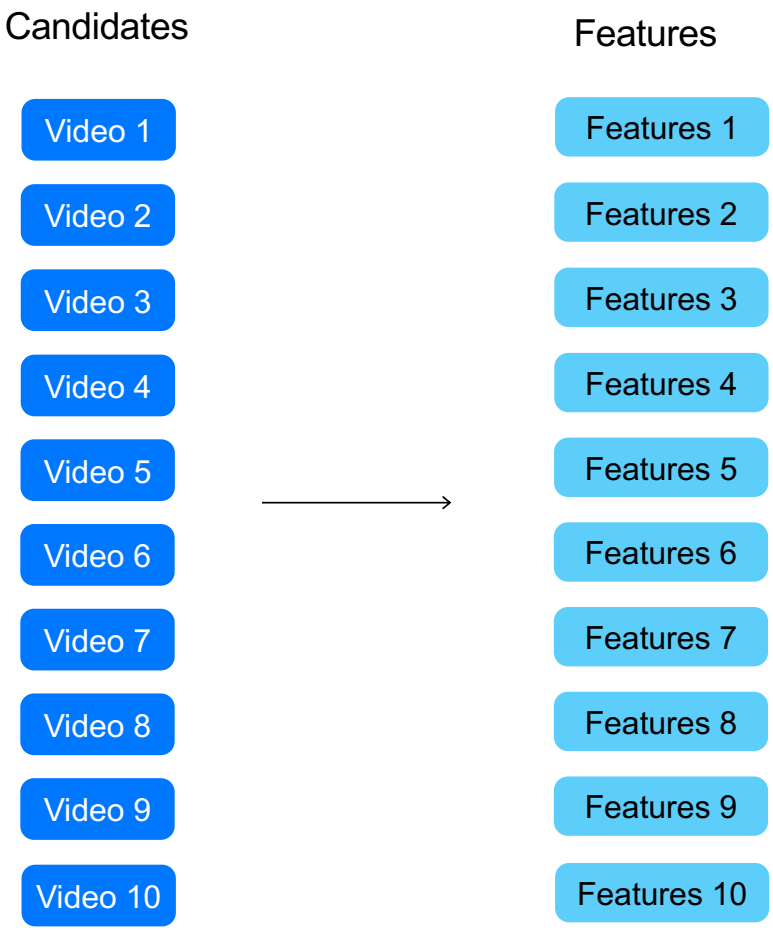
Video 7

Video 8

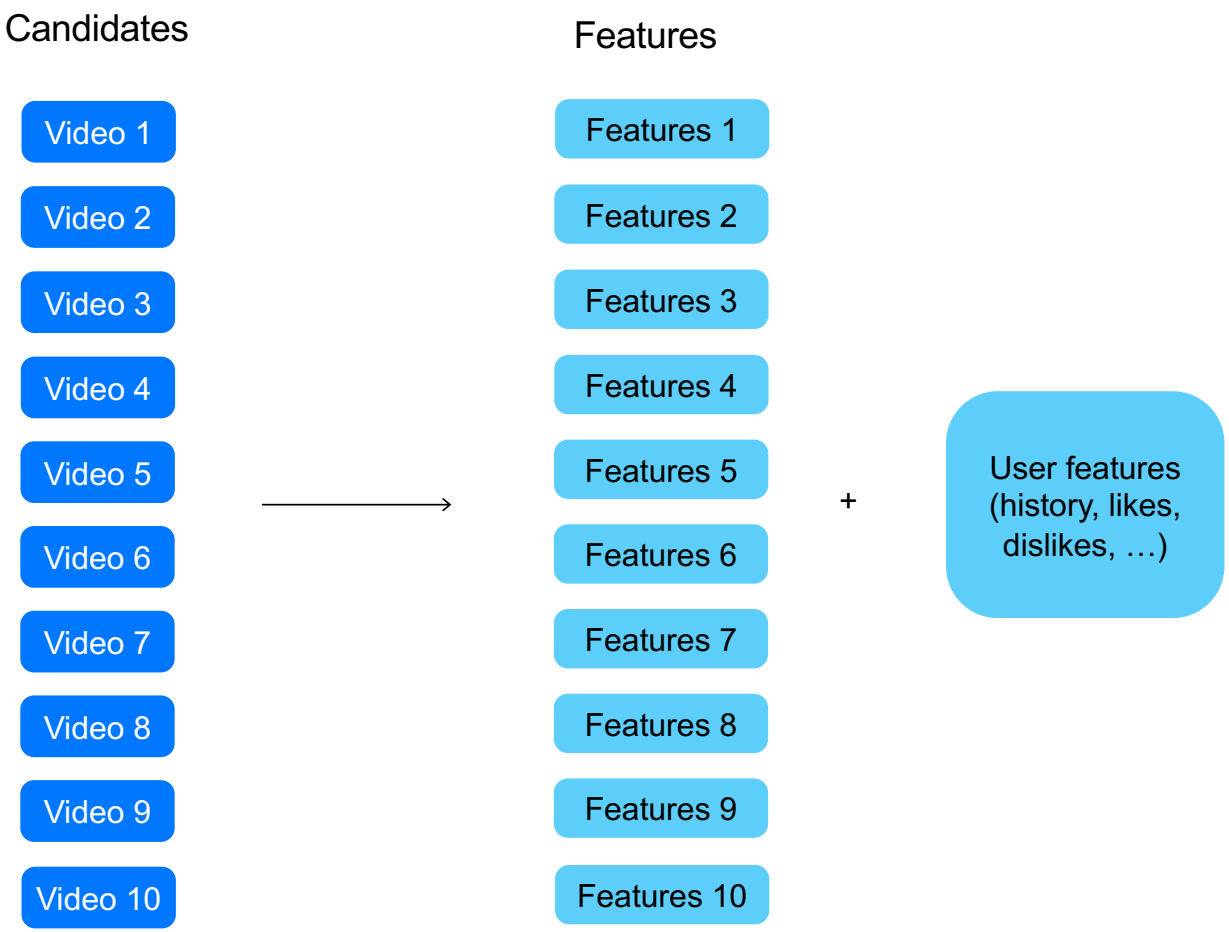
Video 9

Video 10

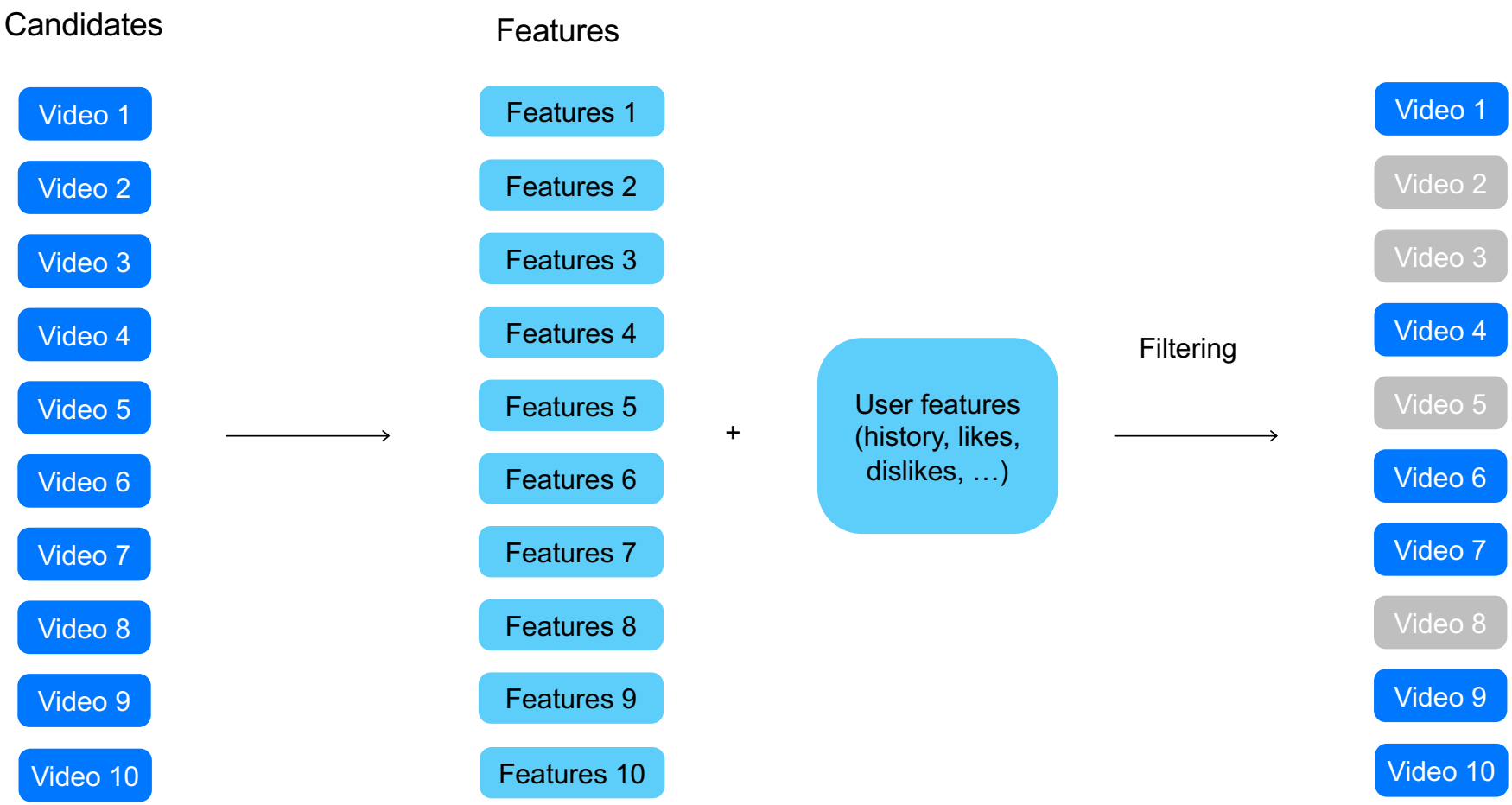
Feed


Candidates
Features
Video 1
Video 2
Video 3
Video 4
Video 5
Video 6
Video 7
Video 8
Video 9
Video 10
Features 1
Features 2
Features 3
Features 4
Features 5
Features 6
Features 7
Features 8
Features 9
Features 10

Feed


Candidates
Features
Video 1
Video 2
Video 3
Video 4
Video 5
Video 6
Video 7
Video 8
Video 9
Video 10
Features 1
Features 2
Features 3
Features 4
Features 5
Features 6
Features 7
Features 8
Features 9
Features 10
+
User features (history, likes, dislikes, …)

Feed


Candidates
Video 1
Video 2
Video 3
Video 4
Video 5
Video 6
Video 7
Video 8
Video 9
Video 10
Features
Features 1
Features 2
Features 3
Features 4
Features 5
Features 6
Features 7
Features 8
Features 9
Features 10
+
User features (history, likes, dislikes, …)
Filtering
Video 1
Video 2
Video 3
Video 4
Video 5
Video 6
Video 7
Video 8
Video 9
Video 10

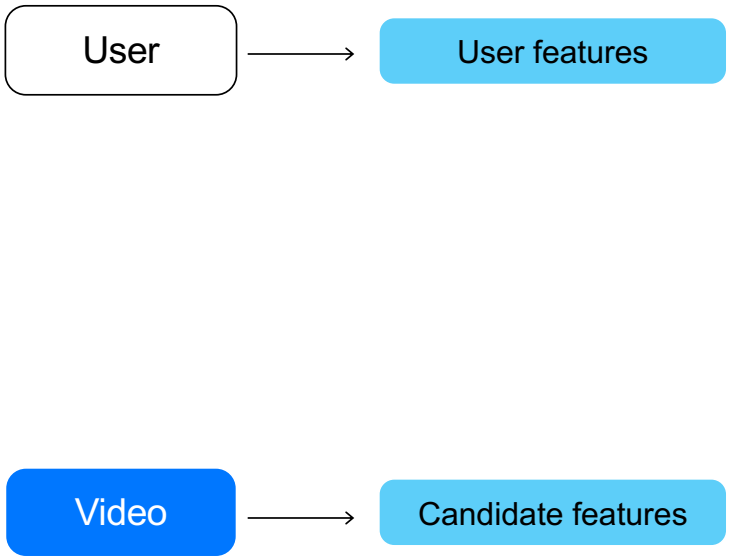
Feed

Feed

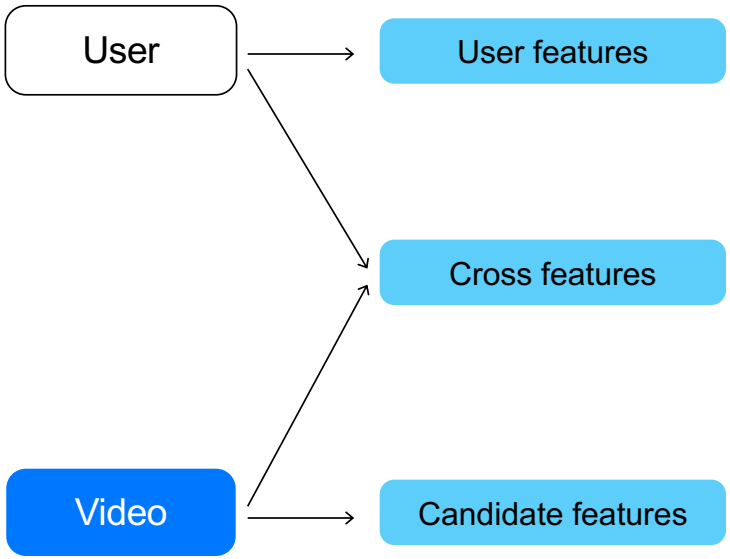
User

Video

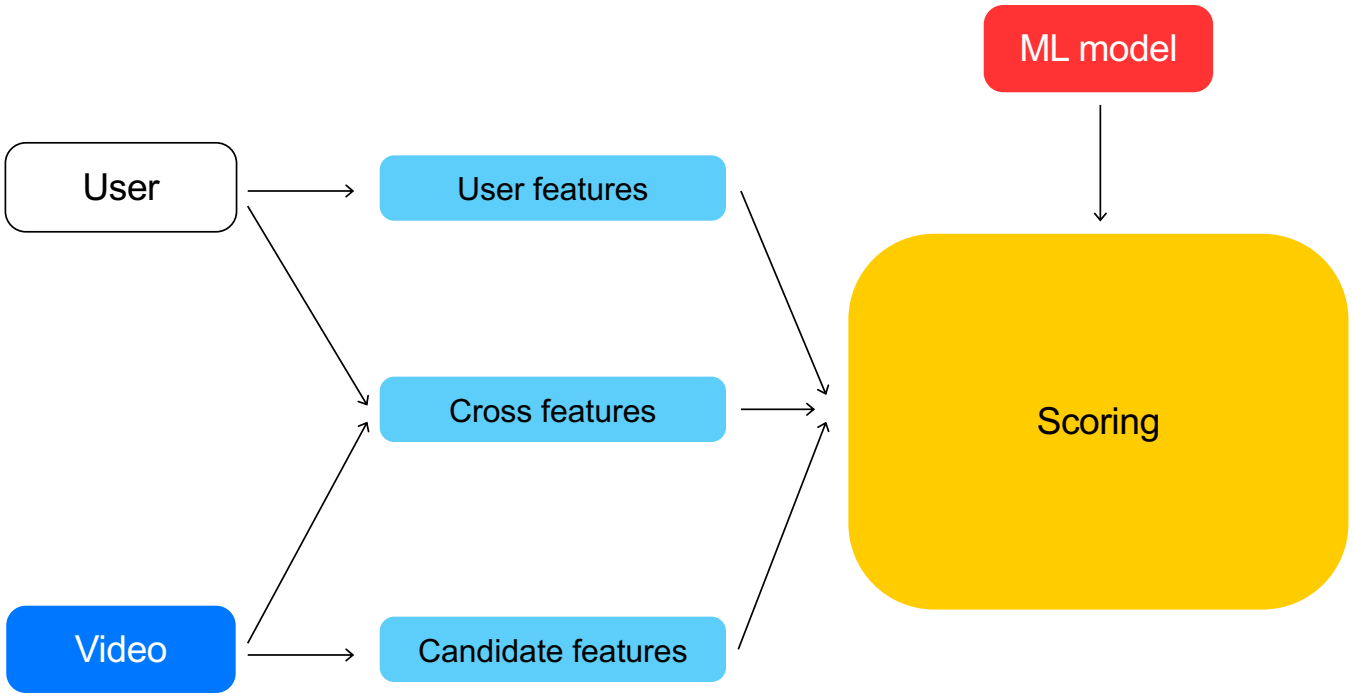
Feed



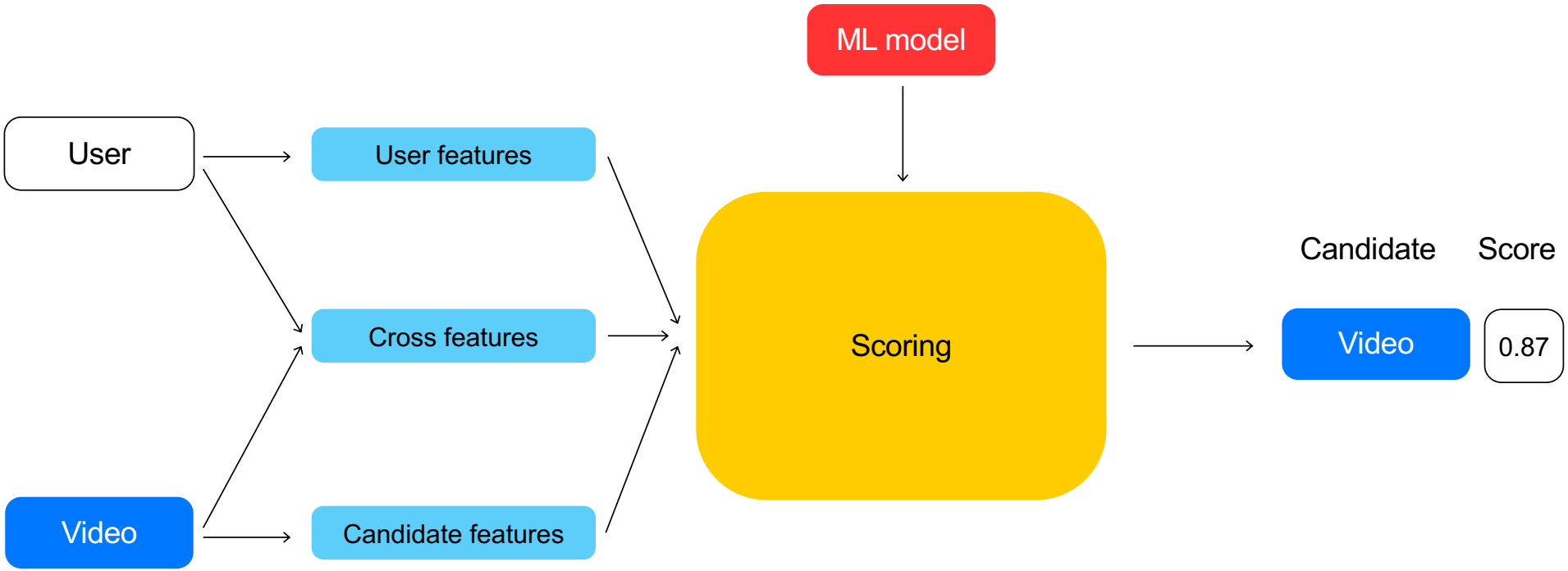
Feed


User
User features
Cross features
Video
Candidate features

Feed


ML model
User
User features
Cross features
Scoring
Video
Candidate features

Feed


ML model
User
User features
Cross features
Video
Candidate features
Scoring
Candidate
Score
Video
0.87

Feed

Feed

Video 1

Video 2

Video 3

Video 4

Video 5

Video 6

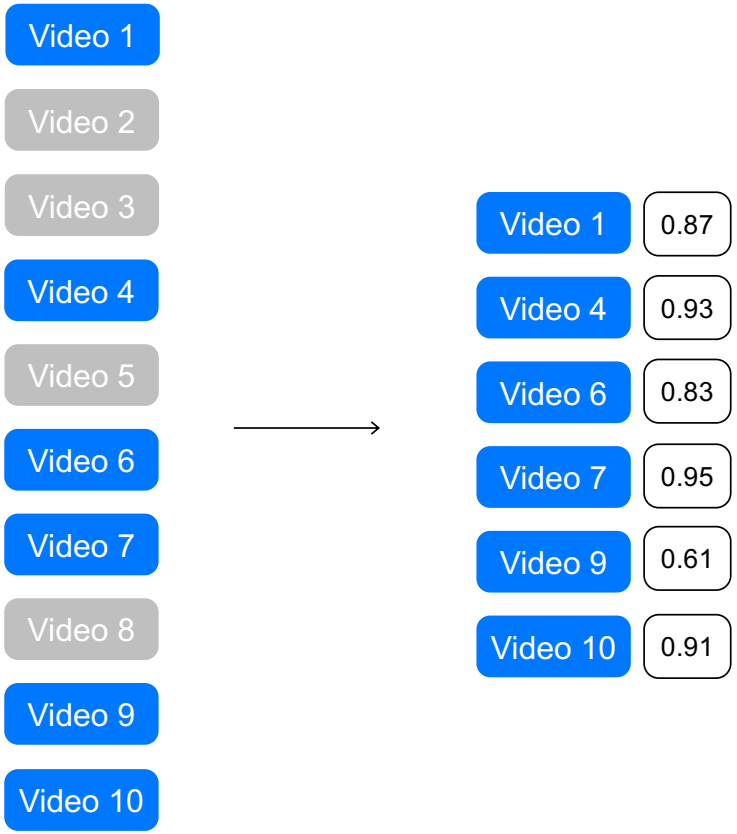
Video 7

Video 8

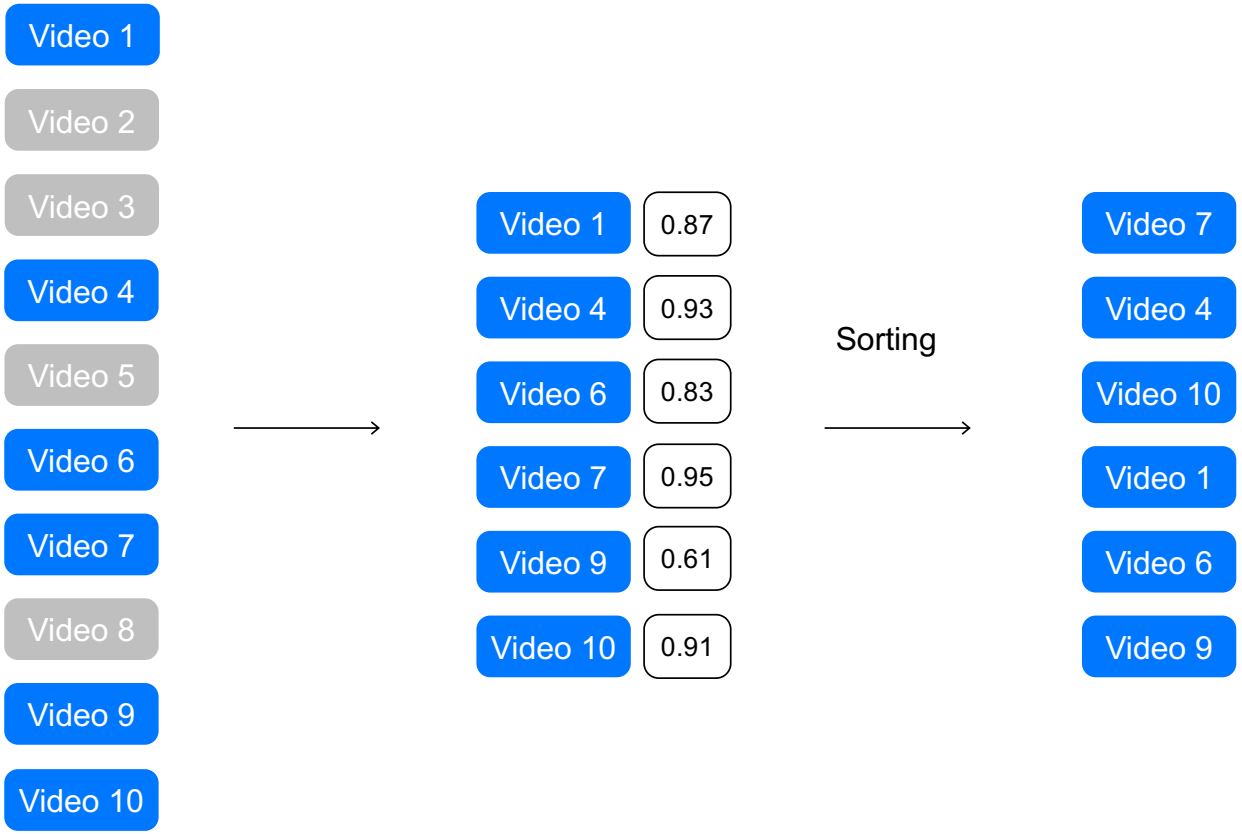
Video 9

Video 10

Feed


Video 1
Video 2
Video 3
Video 4
Video 5
Video 6
Video 7
Video 8
Video 9
Video 10
Video 1
0.87
Video 4
0.93
Video 6
0.83
Video 7
0.95
Video 9
0.61
Video 10
0.91

Feed


Video 1
Video 2
Video 3
Video 4
Video 5
Video 6
Video 7
Video 8
Video 9
Video 10
Video 1
0.87
Video 4
0.93
Video 6
0.83
Video 7
0.95
Video 9
0.61
Video 10
0.91
Sorting
Video 7
Video 4
Video 10
Video 1
Video 6
Video 9

03

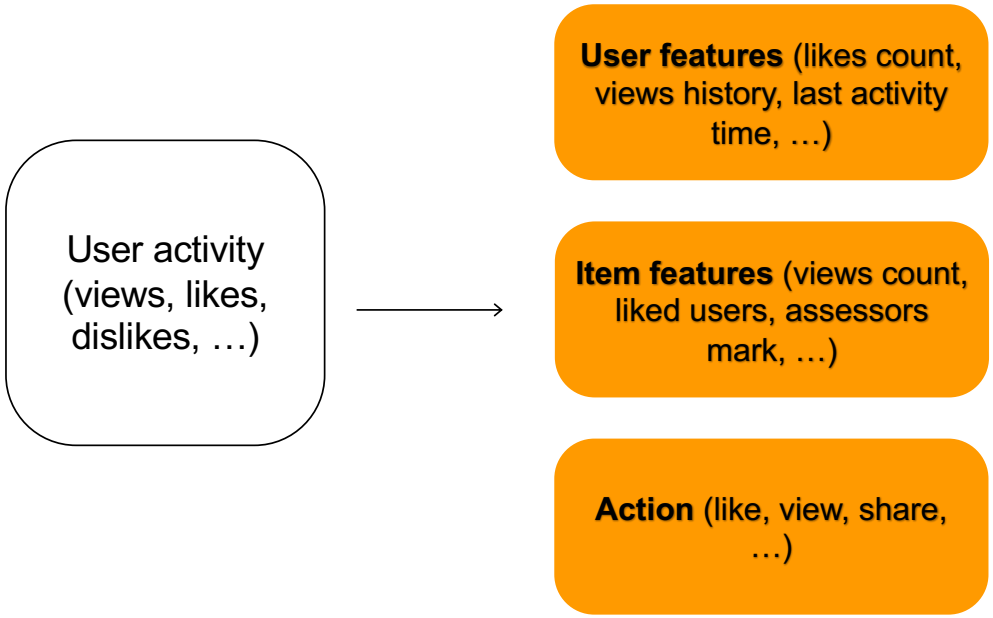
Training

Training

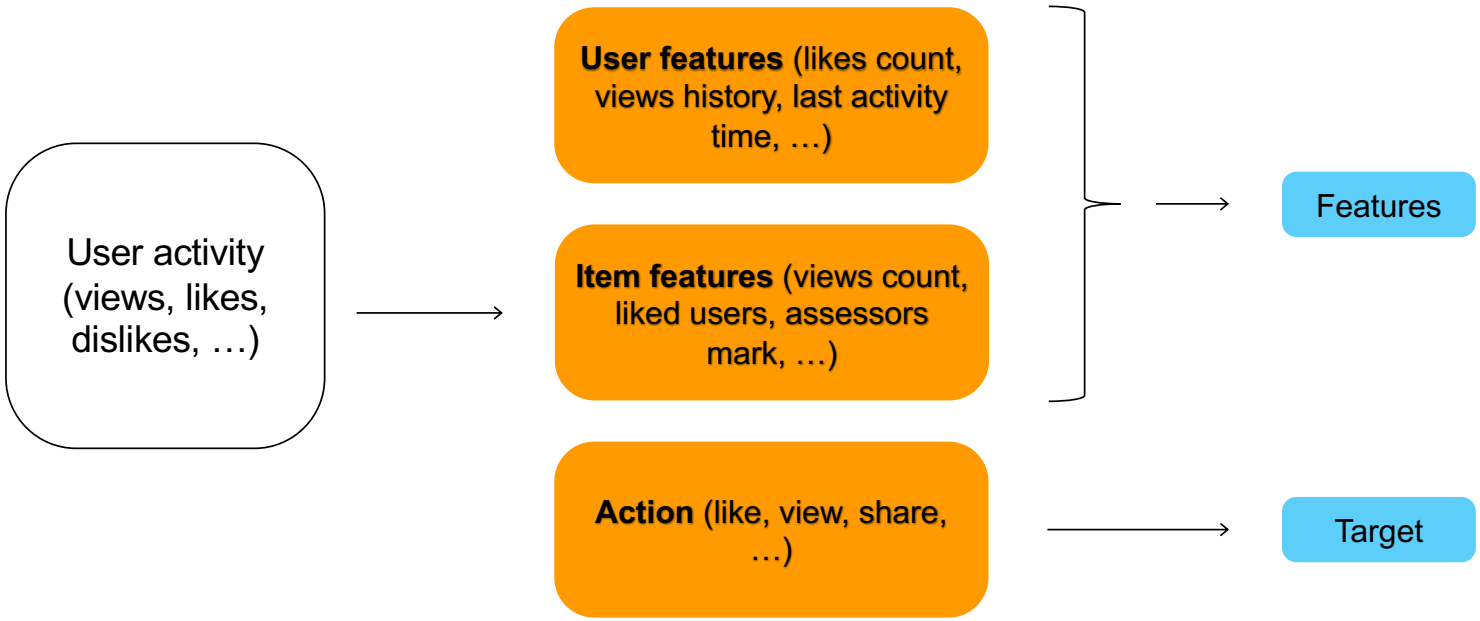
Training

User activity (views, likes, dislikes, …)

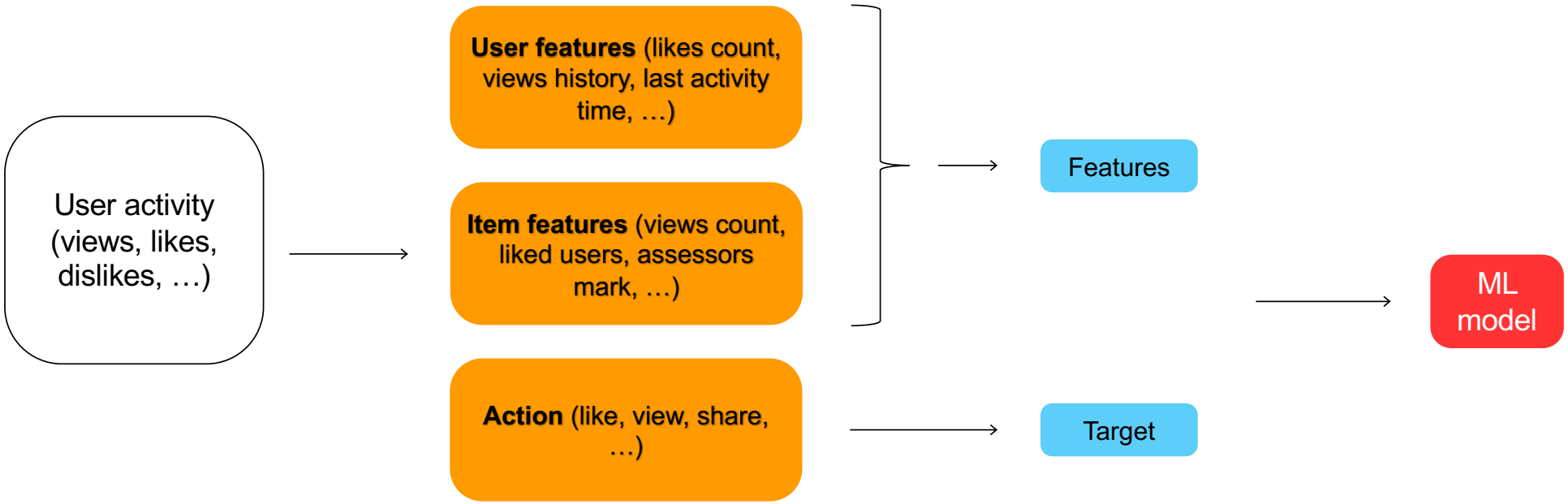
Training


User activity (views, likes, dislikes, …)
User features (likes count, views history, last activity time, …)
Item features (views count, liked users, assessors mark, …)
Action (like, view, share, …)

Training


User activity (views, likes, dislikes, …)
User features (likes count, views history, last activity time, …)
Item features (views count, liked users, assessors mark, …)
Action (like, view, share, …)
Features
Target

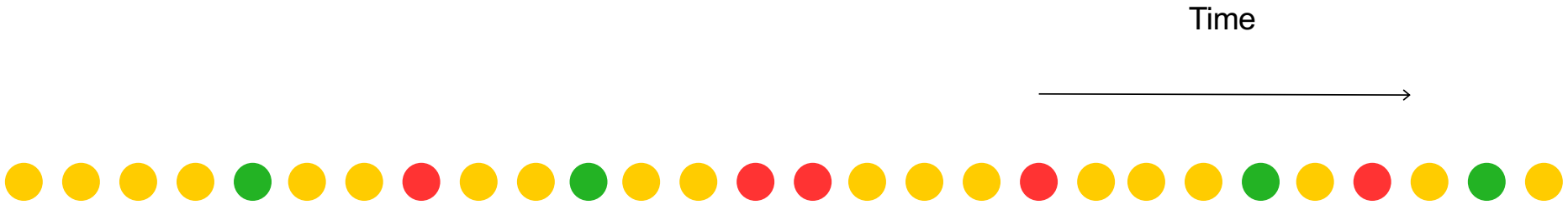
Training


User activity (views, likes, dislikes, …)
User features (likes count, views history, last activity time, …)
Item features (views count, liked users, assessors mark, …)
Action (like, view, share, …)
Features
Target
ML model

Training

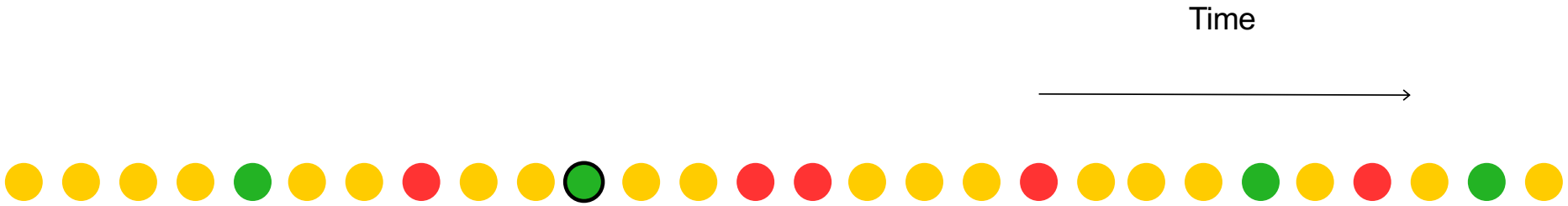
Training

- some action
- good action (like, share, long view, …)
- bad action (dislike, short view, …)



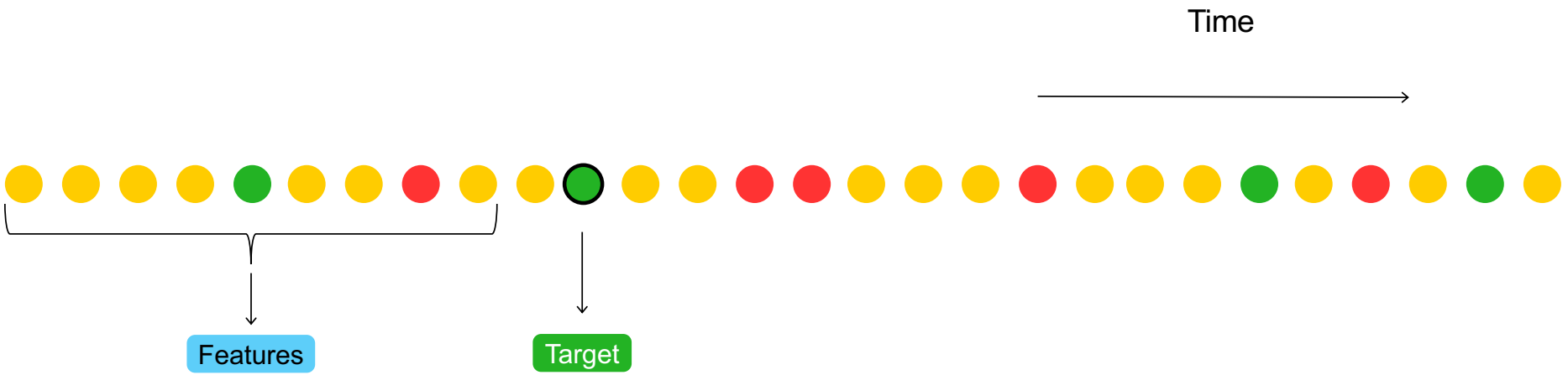
Training

- some action
- good action (like, share, long view, …)
- bad action (dislike, short view, …)



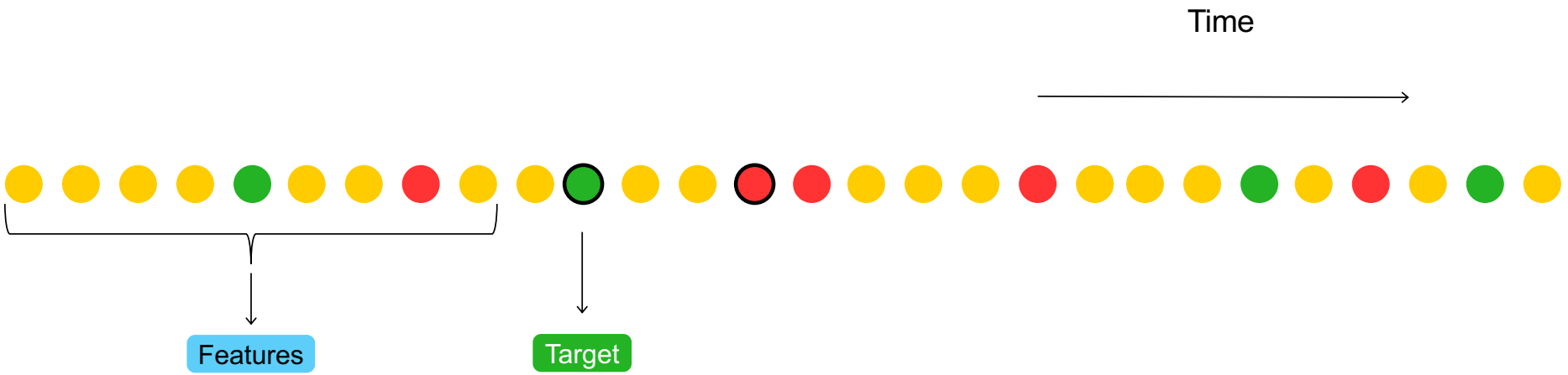
Training

- some action
- good action (like, share, long view, …)
- bad action (dislike, short view, …)



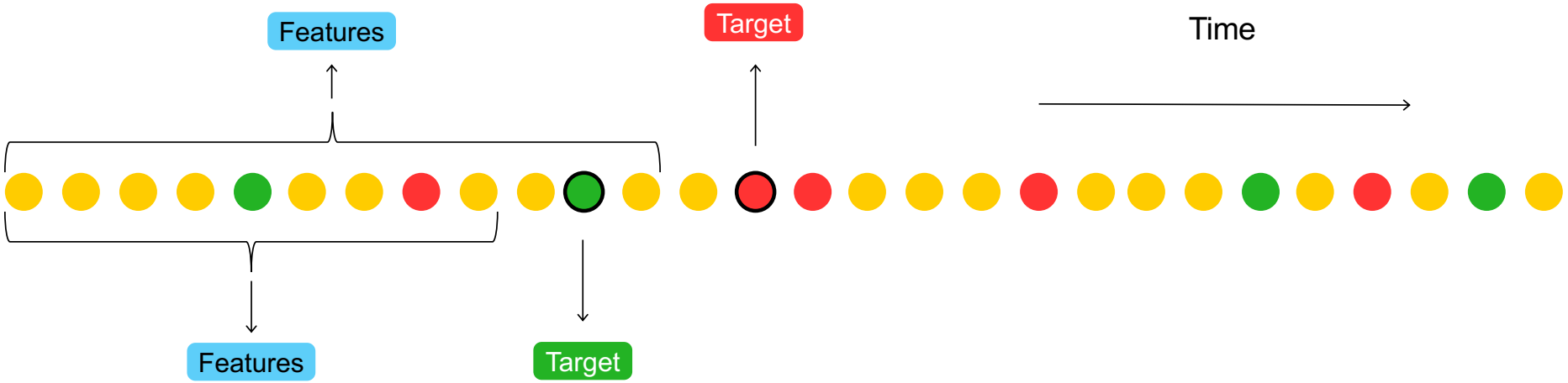
Training

- some action
- good action (like, share, long view, …)
- bad action (dislike, short view, …)



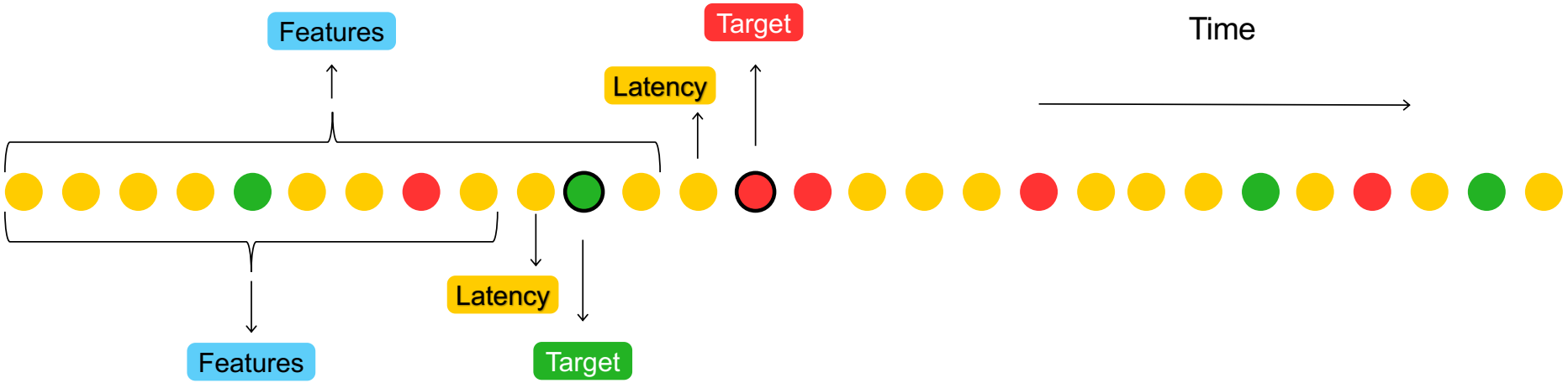
Training

- some action
- good action (like, share, long view, …)
- bad action (dislike, short view, …)



Training

- some action
- good action (like, share, long view, …)
- bad action (dislike, short view, …)



Training

Training

Metrics

- NDCG
- QueryAverage:top=N

Training

Metrics

- NDCG
- QueryAverage:top=N

Loss functions

- QuerySoftMax
- YetiRank

Training

Metrics

- NDCG
- QueryAverage:top=N

Loss functions

- QuerySoftMax
- YetiRank

https://catboost.ai/docs/concepts/loss-functions-ranking.html