



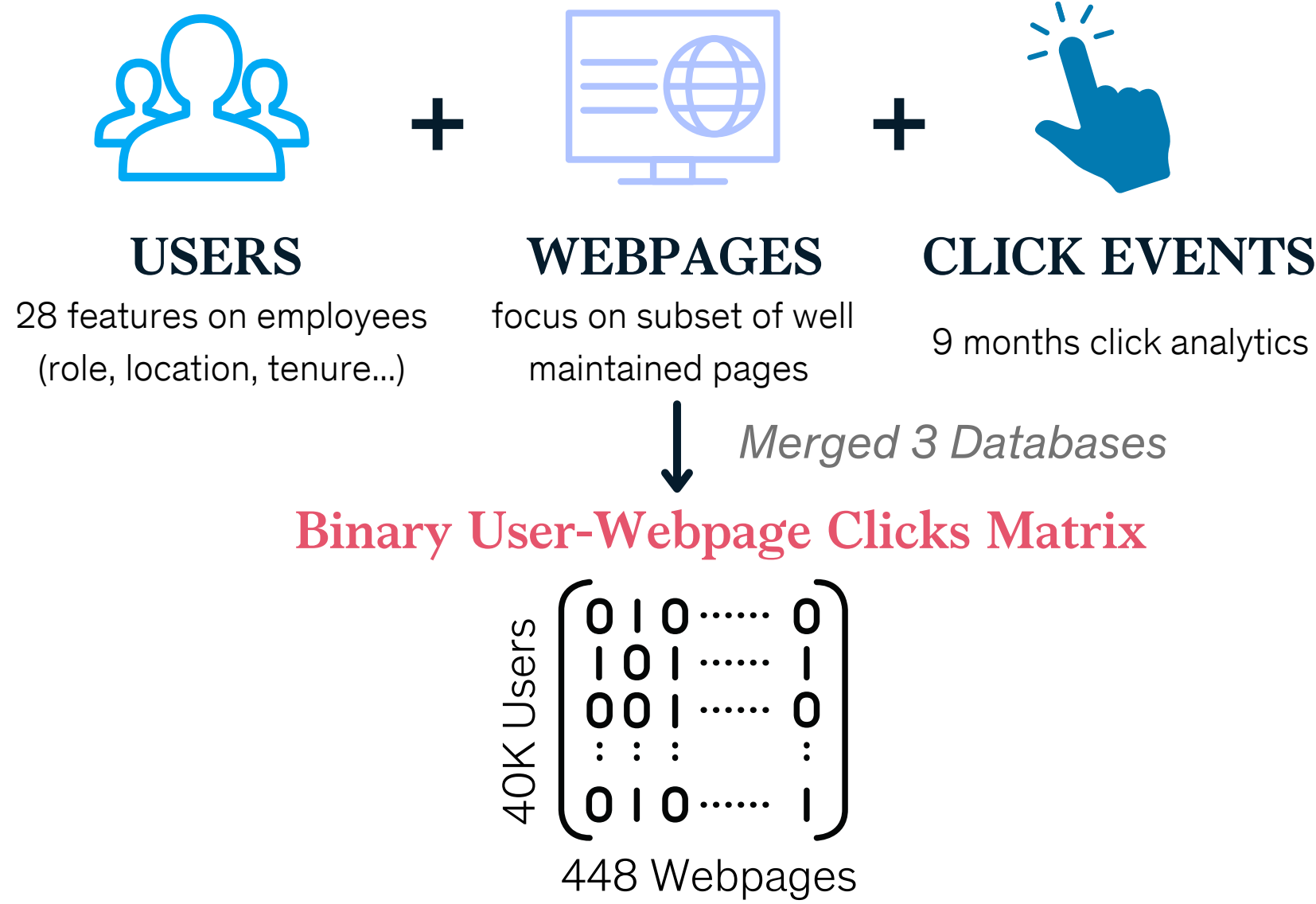
I. PROBLEM STATEMENT

A consulting firm's Global Intranet is a resource that is used by employees to have easy access to tools, information and expertise including firm benefits, learning portals, etc.

The central objective of this project is to **develop a personalized intranet recommendation system** to

- increase user engagement
- offer discoverability to less popular intranet webpages

II. DATA PROCESSING & MATRIX FORMULATION

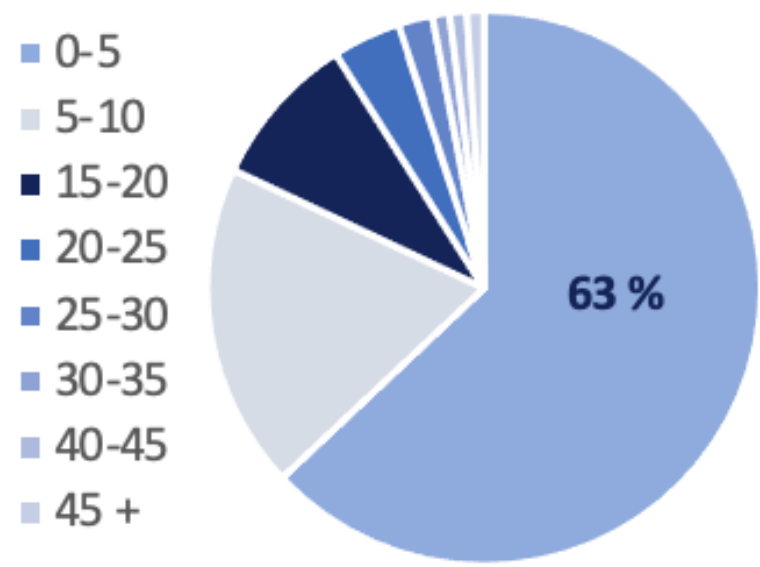


DATA LIMITATIONS

- Matrix Sparsity**: only 1.6% of 16M interactions are non-zero
- Implicit Feedback**: Frequency of clicks doesn't imply more usefulness
- Not Visited (0) ≠ Not useful (pages were not presented)

III. EXPLORATORY DATA ANALYSIS

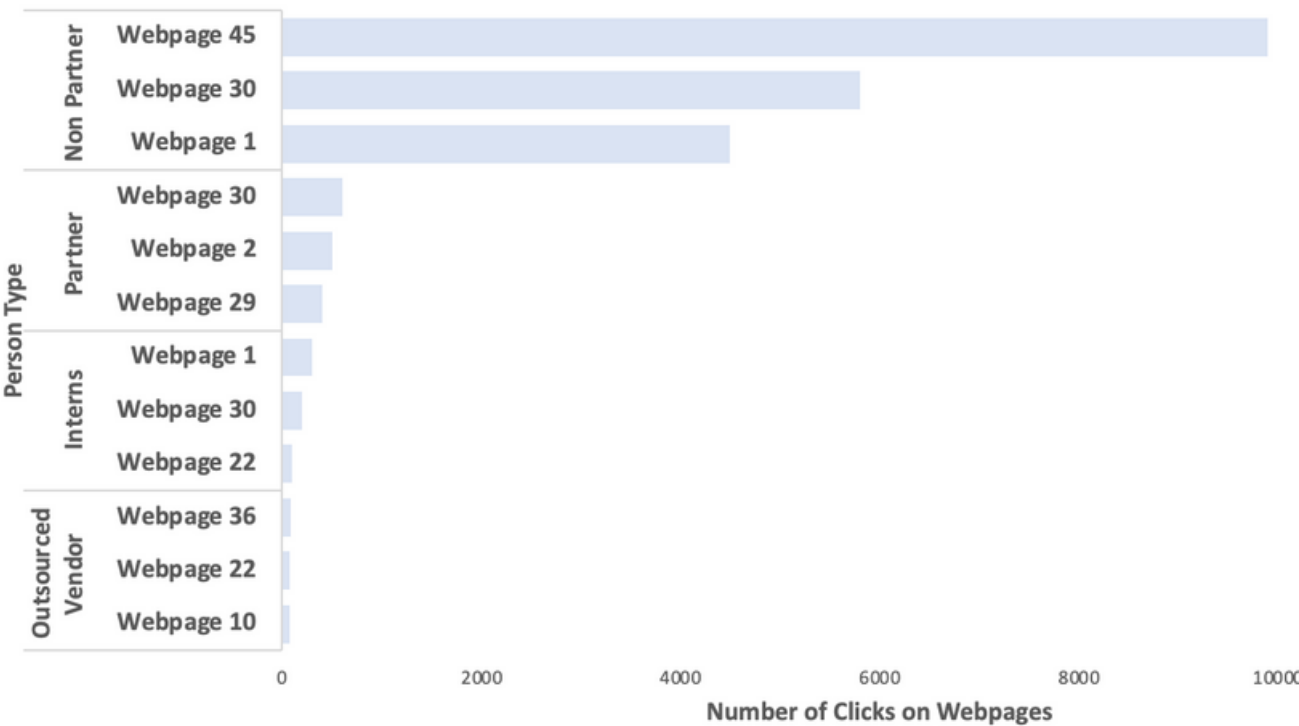
Users' Clicks Distribution



Low Activity

63% of users have a total of <5 clicks --> motivated binary modeling

Visited Content Per Person Type

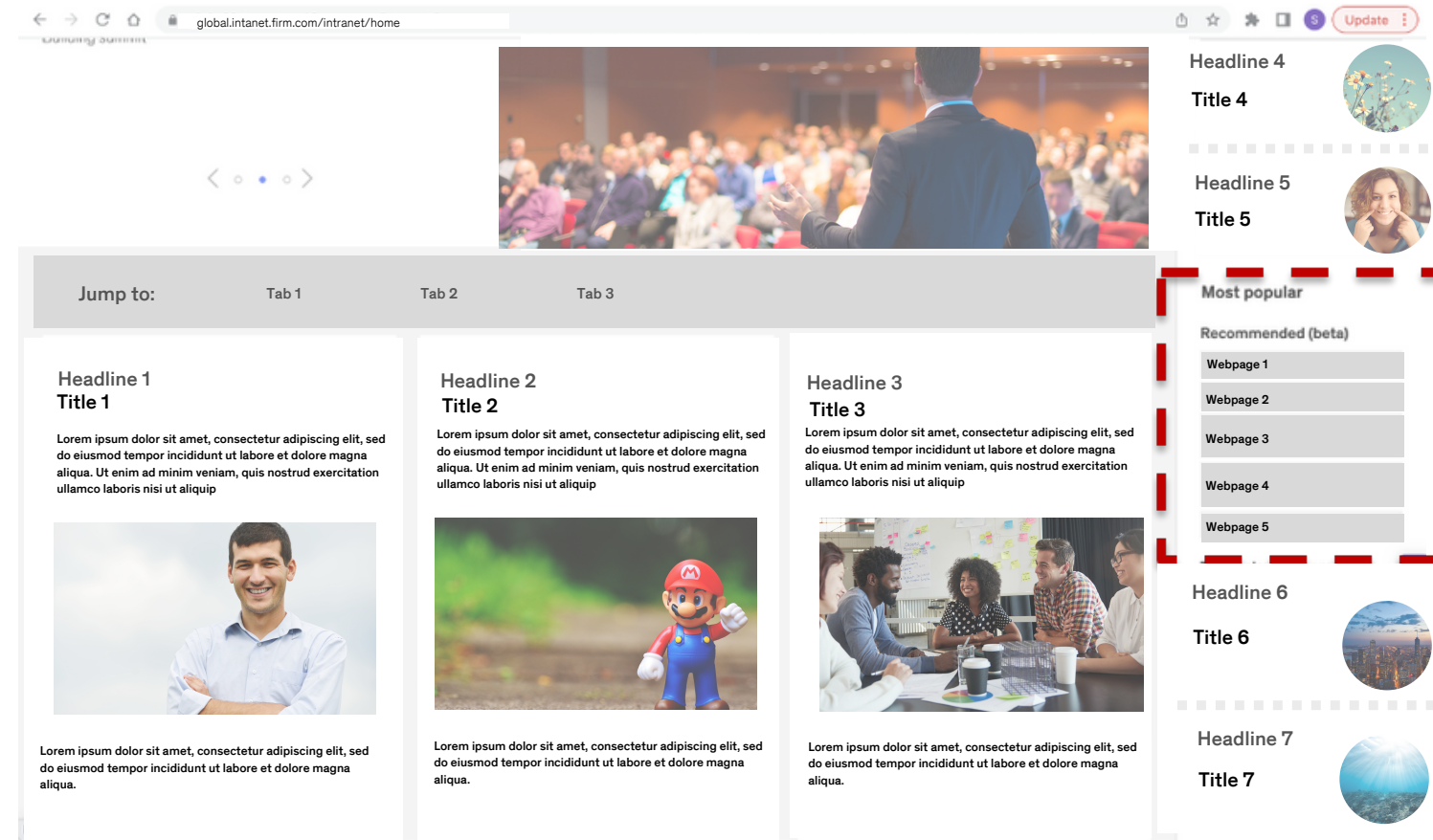


Assured presence of signal

Motivated baseline creation

IV. BASELINE CREATION & DEPLOYMENT

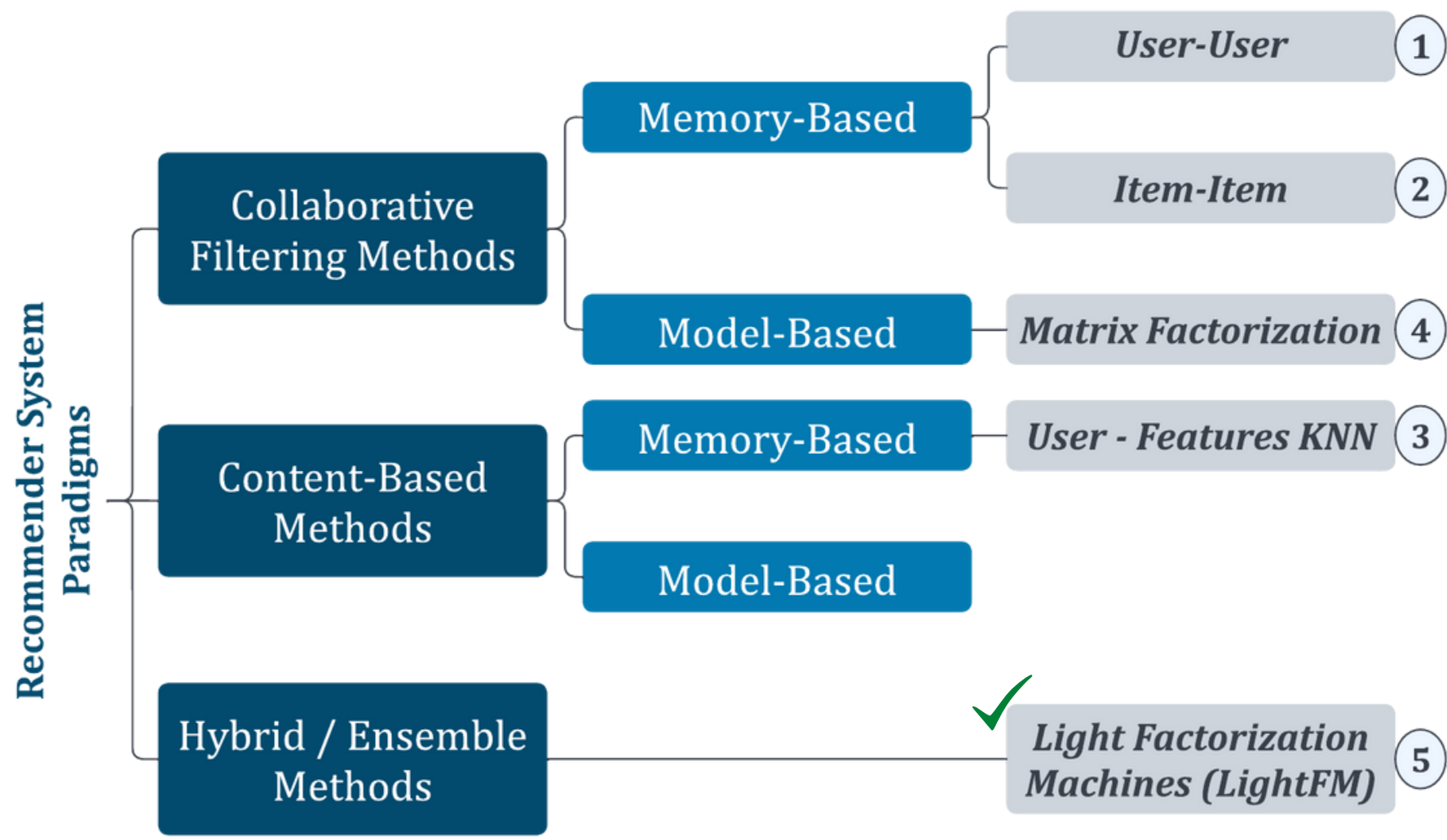
To act as an initial assessment point to measure the performance of our recommender system models, a non-machine learning baseline was created and deployed



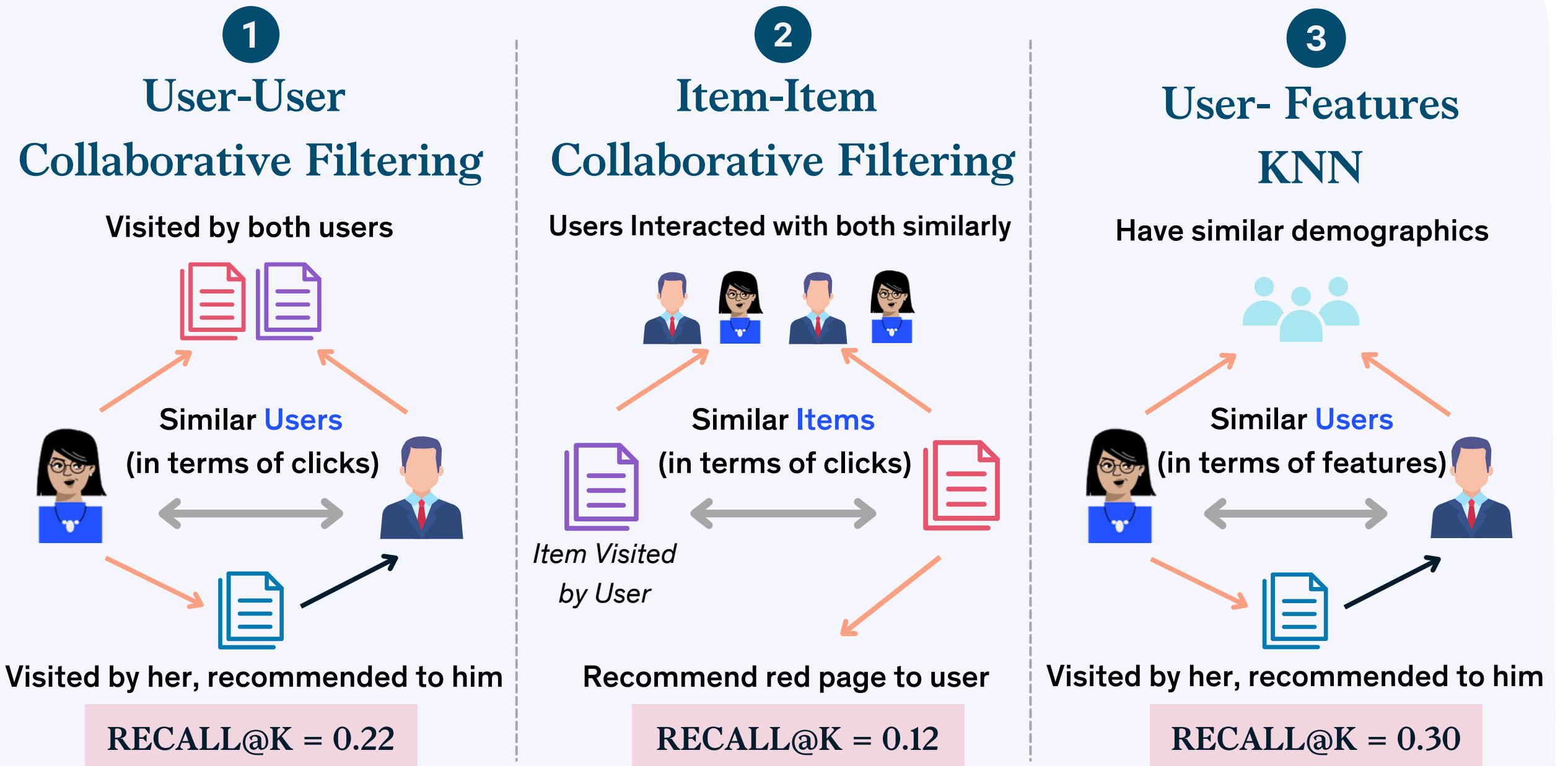
- Top 5 Most Visited Pages
- Per Person Type
- Per Office Location
- RECALL@K = 0.21

V. RECOMMENDER SYSTEM MODELING

Five Candidate Models



Models 1 - 3 are based on KNN and differ in the similarity metric used



Deep-Dive On Chosen Model - LightFM [1]

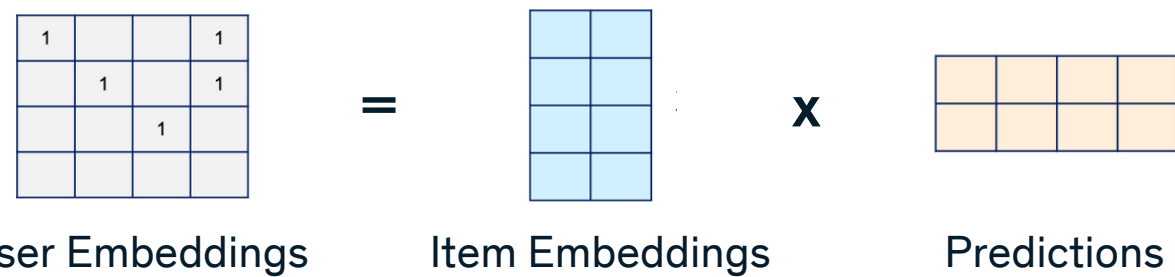
- Leverages clicks + features
- Tackles cold start for new and inactive users
- Ensemble nature deals well with sparsity and implicit feedback
- Highest Recall@K

STEP 1: Incorporating Features in Embeddings

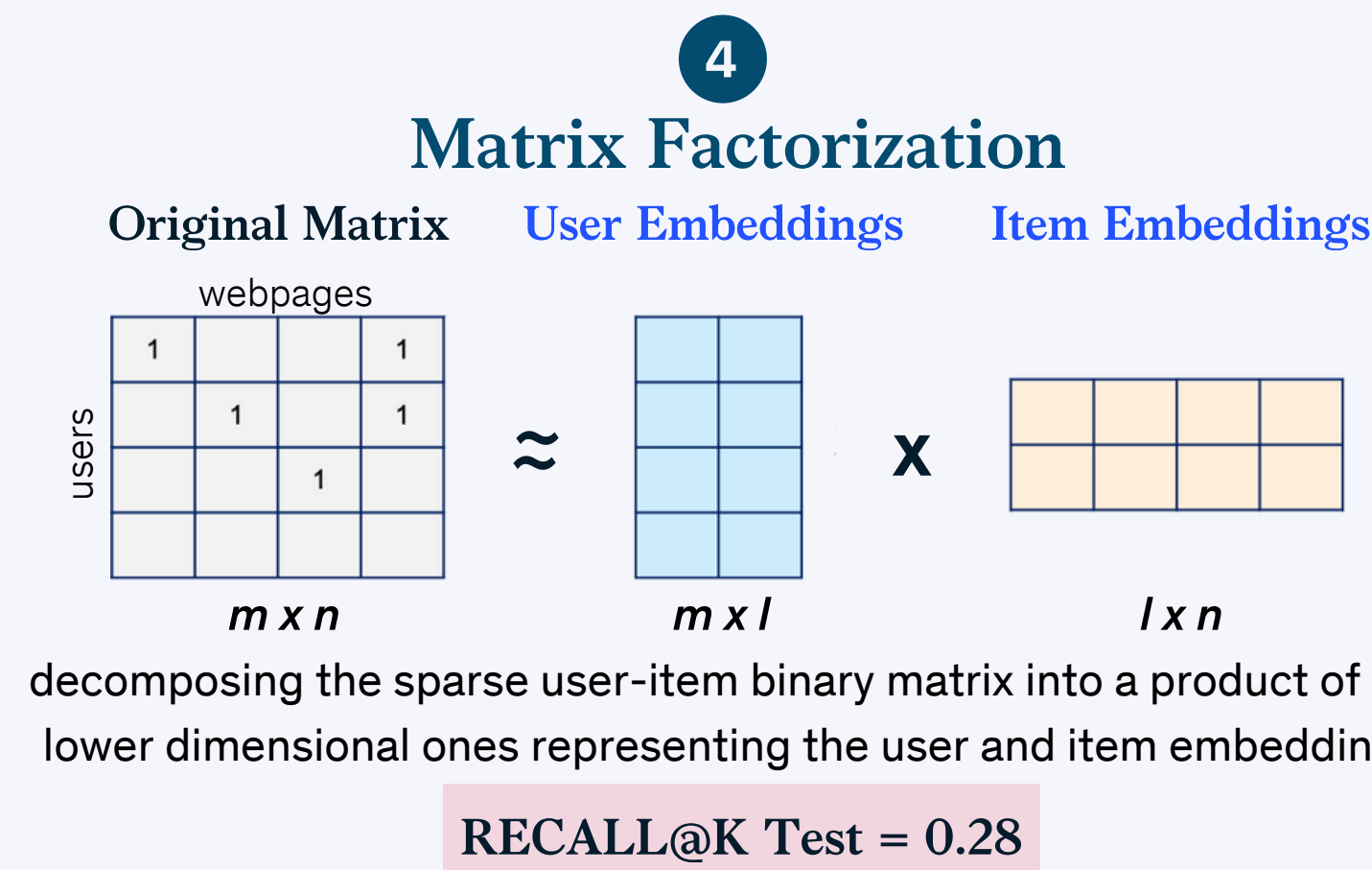
Illustration on subset of user features - the same is done for item features

	CSP	ESP	FSP		Consulting	Operations		Consulting	Operations
User 1	1	0	0	*	0.9	0.1	=	0.9	0.1
User 2	0	0	1		0.2	0.8		0.1	0.9
User 3	0	1	0		0.1	0.9		0.2	0.8

STEP 2: Matrix Factorization



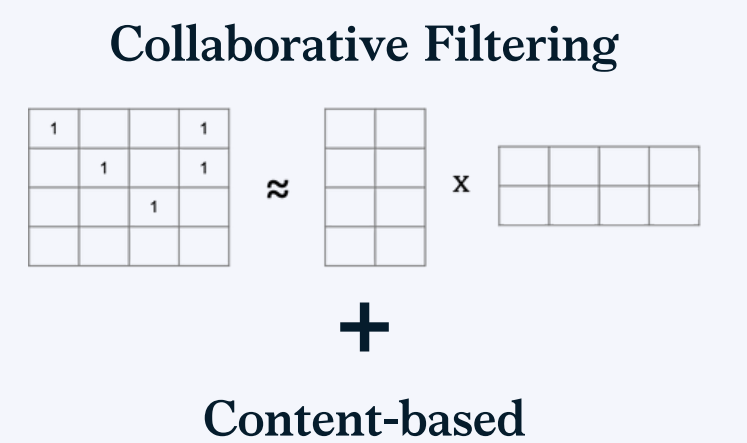
[1] Kula, Maciej, Metadata Embeddings for User and Item Cold-start Recommendations, 07 2015.



decomposing the sparse user-item binary matrix into a product of two lower dimensional ones representing the user and item embeddings

RECALL@K Test = 0.28

LightFM



RECALL@K Test = 0.34

Recall@K (True Positive Rate@K)

Out of the total number of webpages that the model gives (k=5), how many has the user actually visited

VI. RESULTS AND IMPACT

Baseline Productionalized

5K Clicks Per Week

LightFM Model Performs 80% better than baseline