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

- USERS (28 features on employees: role, location, tenure,..)
- WEBPAGES (focus on subset of well-maintained pages)
- CLICK EVENTS (9 months click analytics)

Merged 3 Databases:
- Binary User-Webpage Clicks Matrix

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:
- 63% of users have a total of <5 clicks

User Feature Matrix:
- Low Activity 63% of users have a total of <5 clicks --> motivated binary modeling

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.

V. RECOMMENDER SYSTEM MODELING

Five Candidate Models:
- Collaborative Filtering Methods
- Content-Based Methods
- Memory-Based Methods
- Hybrid/Ensemble Methods
- LightFaction Machines (LightFM)

Deep-Dive On Chosen Model - LightFM

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

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

- User-User Collaborative Filtering
- Item-Item Collaborative Filtering
- User-Features KNN

User-Features KNN

Have similar demographics

V. RECOMMENDER SYSTEM MODELING

Five Candidate Models:
- Collaborative Filtering Methods
- Content-Based Methods
- Memory-Based Methods
- Hybrid/Ensemble Methods
- LightFaction Machines (LightFM)

Deep-Dive On Chosen Model - LightFM

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

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

- User-User Collaborative Filtering
- Item-Item Collaborative Filtering
- User-Features KNN

User-Features KNN

Have similar demographics

VI. RESULTS AND IMPACT

Baseline-Productionized

5K Clicks Per Week

LightFM Model Performs 80% better than baseline