User-Based Collaborative Filtering

Understanding User Similarity for Personalized Recommendations

MUKESH KUMAR

What is Collaborative Filtering?

Recommendation technique

Filters items based on preferences of similar users

User-Based approach

Groups similar users together

Key assumption

Similar people like similar things

Basic Concept

Find similar users

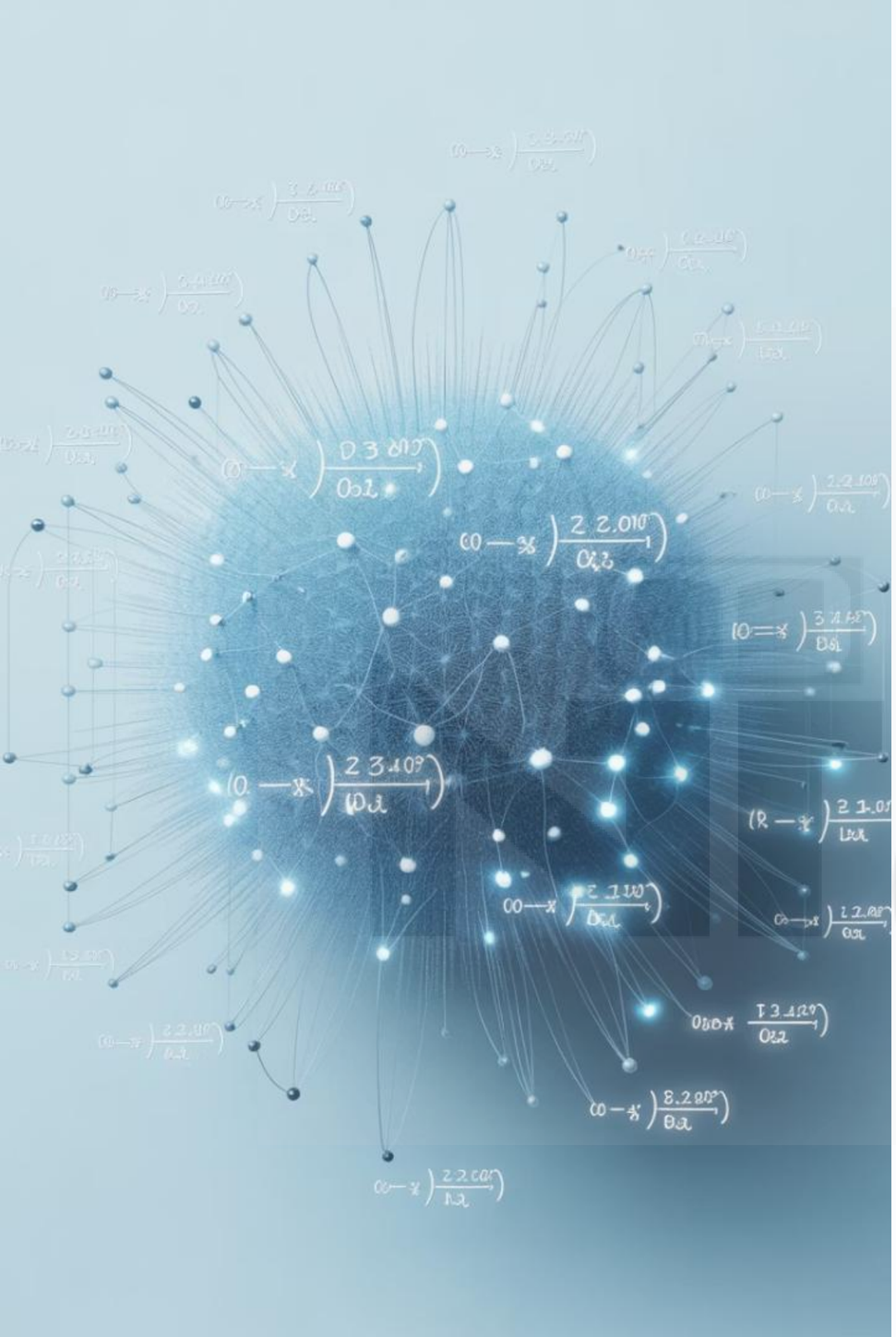
Group people with matching preferences

Identify new items

Find what similar users liked but active user hasn't seen

Generate recommendations

Suggest these new items to active user



User Similarity Methods

Cosine Similarity

Measures angle between user vectors

Pearson Correlation

Considers rating scale differences

Jaccard Similarity

Compares sets of items

User-Item Matrix

	Movie 1	Movie 2	Movie 3	Movie 4
User A	5	3	-	4
User B	4	-	5	3
User C	5	2	-	-

Step-by-Step Process

Create user-item matrix

Map all user interactions with items

Compute similarity

Calculate how similar each user is to others

Find neighbors

Identify top-N most similar users

Aggregate preferences

Collect items liked by similar users

Rank and recommend

Score and suggest new items to user

Example Walkthrough

User A

- Liked: Item 1, Item 3
- Similar to: User B
- Recommendation: Item 5

User B

- Liked: Item 1, Item 3, Item 5
- Similar to: User A

User C

- Liked: Item 2, Item 4
- Not similar to others

Pros and Cons

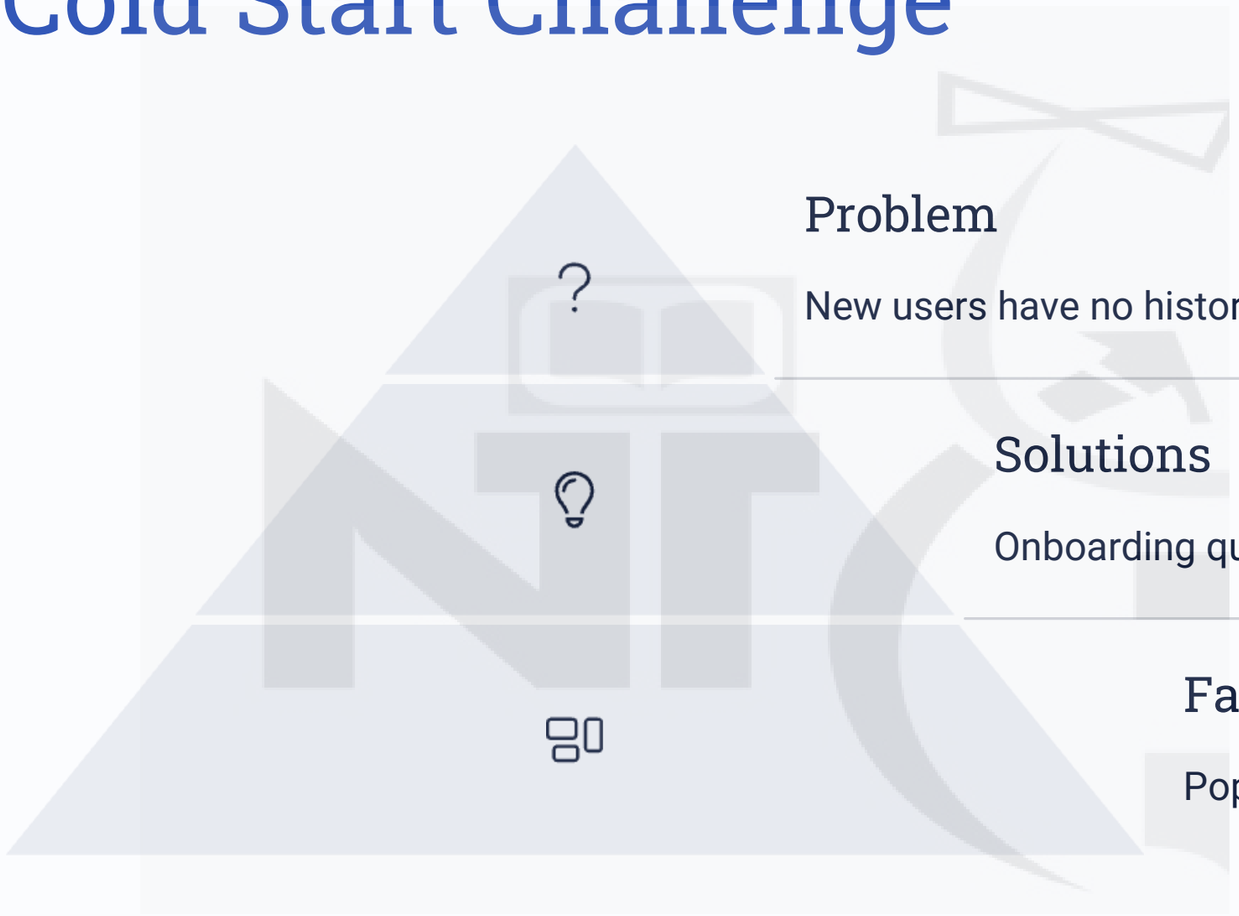
Pros

- Simple implementation
- No content analysis needed
- Discovers unexpected items

Cons

- Scalability issues
- Cold start problem
- Sparsity challenges

Cold Start Challenge



Problem

New users have no history

Solutions

Onboarding questions, demographics, hybrid approaches

Fallbacks

Popular items, trending content

Evaluation Metrics

Accuracy

Precision, Recall, F1-score

Error Rates

RMSE, MAE for rating predictions

User Experience

Coverage, Diversity, Novelty

Implementation Challenges

Data Sparsity

Most user-item matrices are >95% empty

Computational Cost

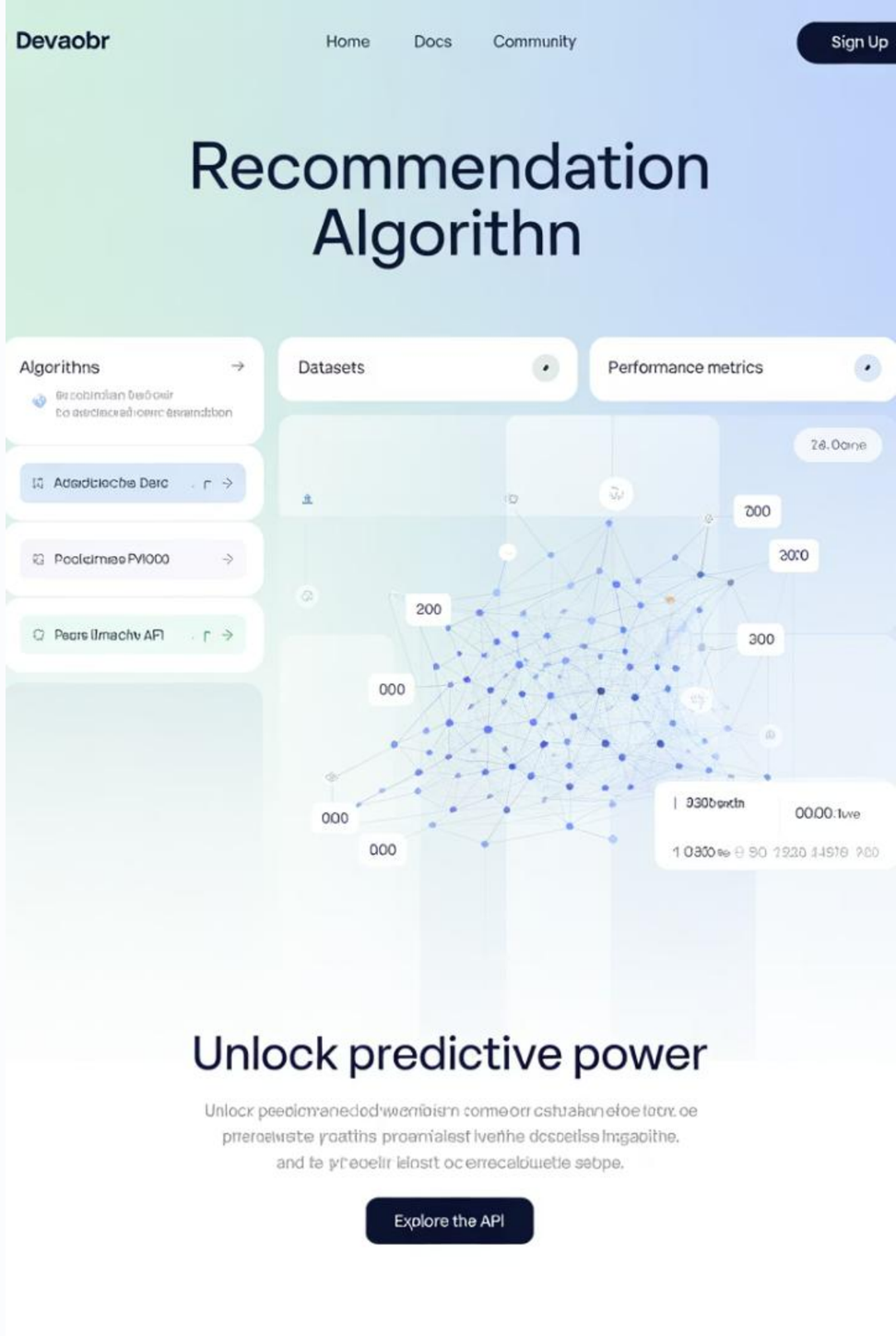
O(n²) complexity as users increase

Real-time Updates

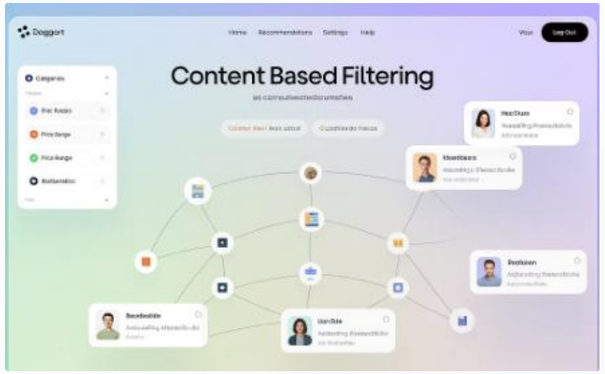
Recalculating similarities is expensive

Privacy Concerns

User preferences reveal sensitive information



Hybrid Approaches

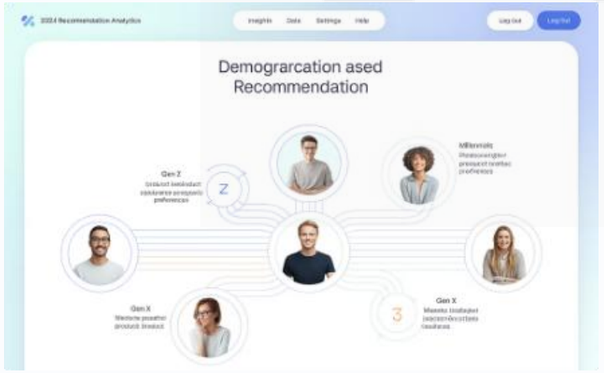


Content + Collaborative

Combines item features with user similarity

Knowledge + Collaborative

Uses domain expertise with collaborative data



Demographic + Collaborative

Groups similar user profiles together