Product Recommendation System

Personalized Recommendations for Loyal and New Customers

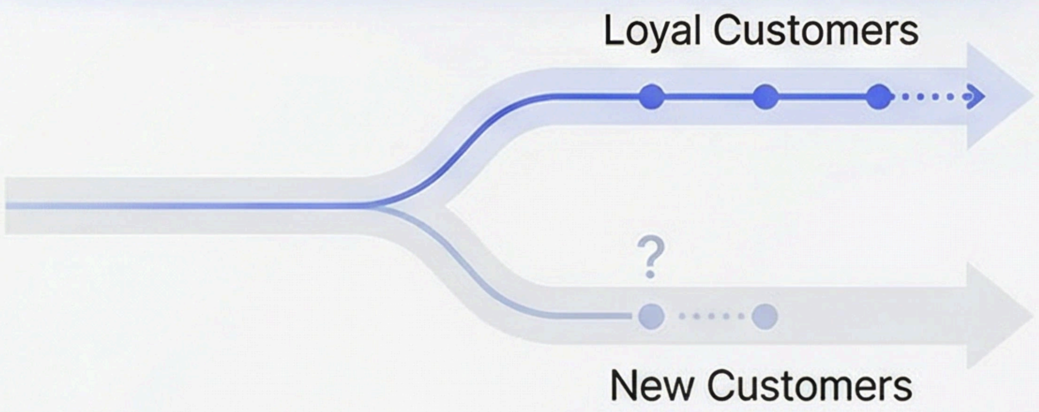
The Challenge: Personalization at Scale

Key Objectives

- Identify 5 relevant products per user
- Handle distinct user segments (Loyal vs. New)
- Process 75,000+ products
- Balance personalization with cold-start

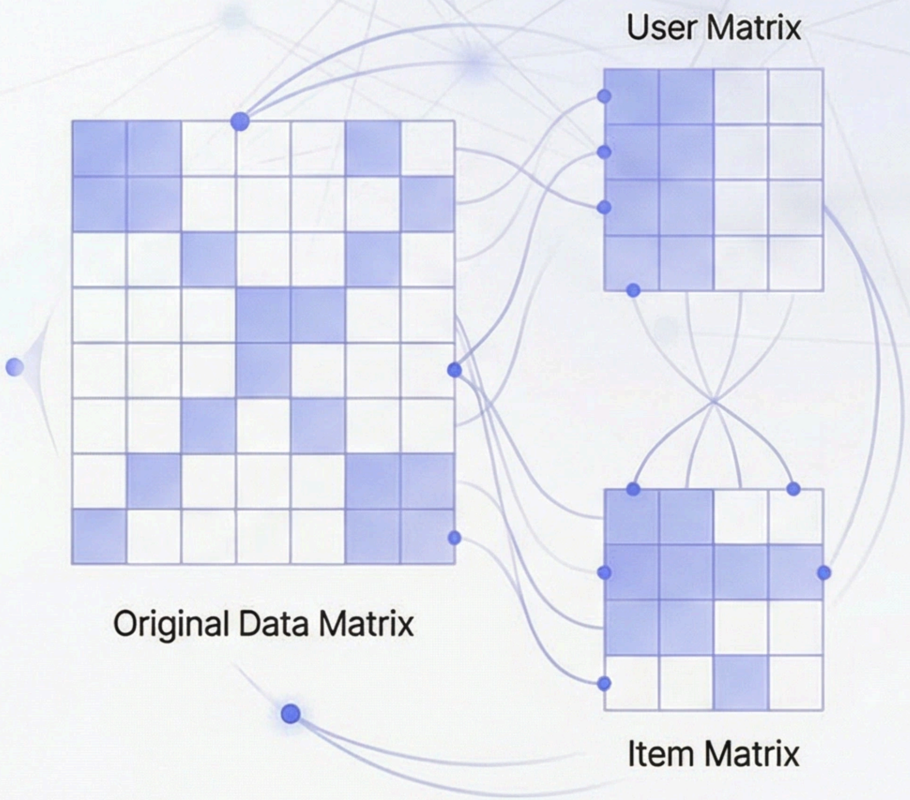
Key Insight

Different users require different approaches—loyal customers have rich history, new customers need alternative strategies.

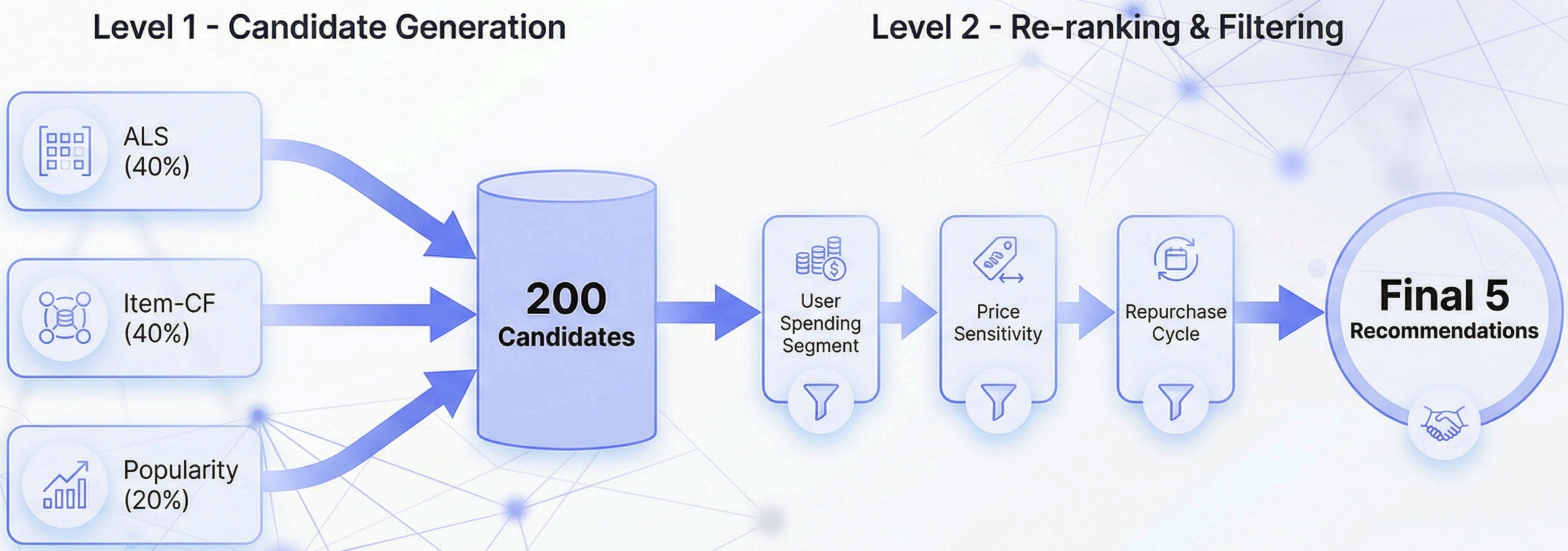


Alternating Least Squares (ALS): The Optimal Choice

- **Effectiveness:** Superior with implicit feedback
- **Scalability:** Efficient for large datasets
- **Implicit Signals:** Uses spending & frequency
- **Latent Factors:** Learns hidden patterns
- **Robustness:** Handles sparse data



Architecture: Candidate Generation + Re-ranking



Understanding User Intent Through Behavior

Spending Ratio Signal

- **item_price / user_avg_item_price**
- Ratio > 1 indicates higher engagement/value perception.

Quantity Ratio Signal

- **units_sold / item_avg_quantity**
- Ratio > 1 indicates strong preference or bulk purchase.

Combined Strength

Weighted 50/50 | Log Transformed | Capped [0.1 - 5.0]

Users reveal true preferences through purchasing behavior.

Intelligent Post-Processing Decisions

Price Sensitivity Adjustment

Soft penalty for items > 1.5x user's avg price. Maintains diversity while respecting budget.

Repurchase Cycle Tracking

Excludes recent purchases (within 50% of cycle). Boosts items due for repurchase (+15% score).

User Segment Filtering

Segments: Small, Low Average, Average, High. Tailors recommendations to spending power.

Revenue Optimization: Upsell Configuration

Current Upsell Strategy

Boost higher-priced items for revenue growth.

Configurable upsell factor (0 = no upsell, 1 = max upsell).

Price boost weight: 20%.

Tunable Parameters

All objectives are tunable via `tuning_config.py`:

- **Upsell factor**: 0.1 (default, adjustable)
- **Price boost weight**: 0.2 (adjustable)
- **Max price ratio**: 2.0x user average (configurable)
- **Model weights**: Adjust ALS/Item-CF/Popularity blend

Future Opportunities

Strategic expansion possibilities:

- **High-margin products**: Recommend premium alternatives at similar price points
- **Private label brands**: Promote store brands with better margins while maintaining value perception
- **Margin-aware recommendations**: Balance revenue optimization with customer satisfaction

Balance revenue optimization with user satisfaction—aggressive upselling may harm long-term engagement.

Ensuring Data Quality for Model Training

Removed 21 Records

Negative units/amounts (Returns/Refunds) removed to prevent model distortion.

Kept 5,600 Zero-Price Items

Promotional items retained to preserve valuable co-purchase signals.

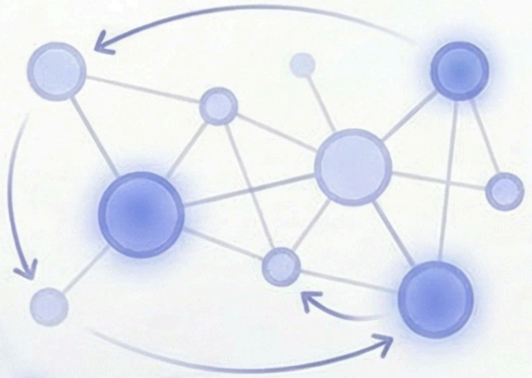
Price Variance Flagged

Extreme fluctuations (e.g., $1.99 to $40) noted for future investigation.

Final Data Quality: 99.97% Retention (77,771 Clean Records)

Item-Item Similarity for Users Without History

Item-Item Collaborative Filtering



- Leverages co-purchase patterns from loyal customers.
- Identifies items frequently bought together.
- Uses spending-weighted similarity for stronger signals.

Popularity Boost & Fallback

- Incorporates 20% popularity signal (purchase count, unique buyers).
- Applies time-based decay to favor recent trends.
- Fallback: Recommend popular items filtered by median price.

Production-Ready Architecture

Backend (FastAPI)

- REST API
- Caching
- Health Checks

Models (Serialized)

- ALS (Alternating Least Squares)
- Item-CF (Item-based Collaborative Filtering)
- User Profiles

Frontend (React)

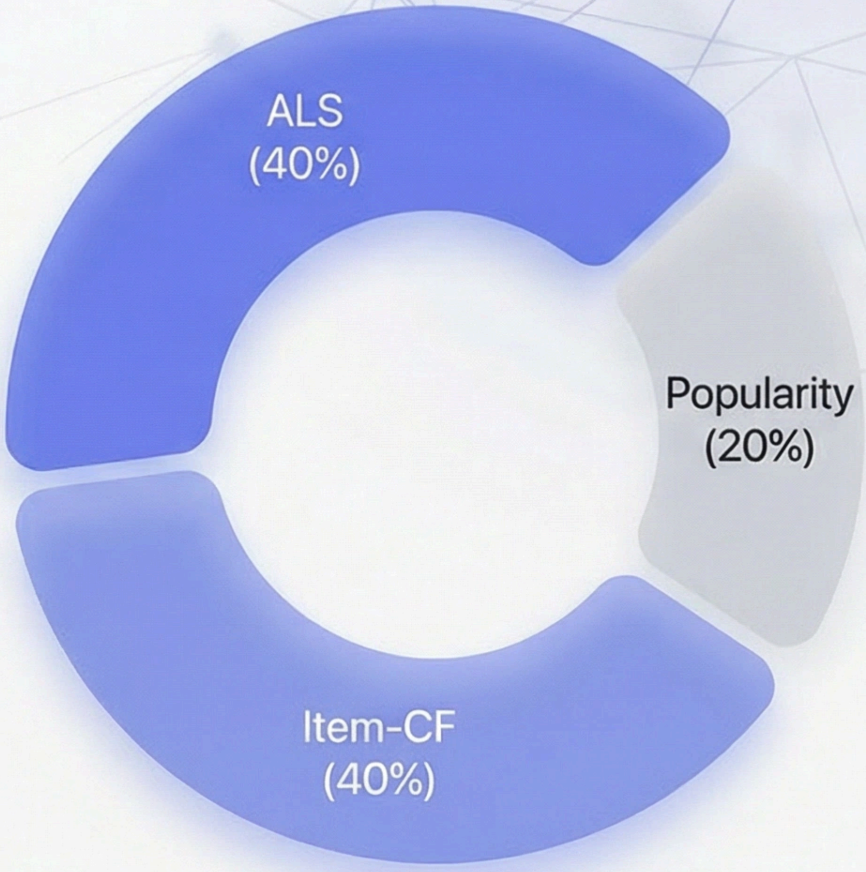
- Model Selection
- Score Visualization

Offline Training

- Serialized Models
- A/B Testing

Combining Strengths: Hybrid Model Approach

- **ALS (40%):** Captures latent user preferences & implicit feedback.
- **Item-CF (40%):** Leverages co-purchase patterns & stability.
- **Popularity (20%):** Ensures diversity & handles cold-start.



Model Selection: Flexible Architecture

ALS (Alternating Least Squares)

- **Best for:** Users with rich purchase history
- **Learns:** Latent user and item factors
- **Output:** Personalized recommendations

Item-Item Collaborative Filtering

- **Best for:** Co-purchase pattern discovery
- **Learns:** Which items are bought together
- **Output:** Stable, interpretable recommendations

Hybrid (40% ALS + 40% Item-CF + 20% Popularity)

- **Best for:** Balanced personalization and diversity
- **Learns:** Combined patterns from all models
- **Output:** Diverse, high-quality recommendations

Popularity-Based

- **Best for:** New users or cold-start scenarios
- **Learns:** Global trending items
- **Output:** Trending recommendations

All objectives are tunable via configuration—choose the model that best matches your business goals.

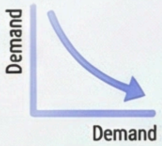
Price Elasticity Analysis: Understanding Purchase Behavior

Key Finding

📊 Analysis of 721 items with variable prices reveals mixed price sensitivity.
📉 Overall correlation (Price vs Units): -0.0525 (weak negative).

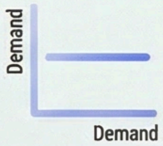
1. **Price Sensitive Items** (296 items)

- **Elasticity:** Negative (e.g., -4.12 to -2.3)
- **Insight:** 'Users buy MORE when price is LOW'
- **Action:** 'Highlight during promotions'



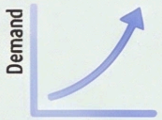
2. **Neutral Items** (319 items)

- **Elasticity:** Near zero
- **Insight:** 'Price has minimal impact'
- **Action:** 'Focus on relevance, not price'



3. **Price Insensitive Items** (106 items)

- **Elasticity:** Positive (e.g., +1.32 to +2.81)
- **Insight:** 'Users buy MORE when price is HIGH (perceived quality?)'
- **Action:** 'Premium positioning strategy'



Important Discovery

Price Variation Finding

Same items show significant price fluctuations (e.g., $1.99 to $40). Likely causes: Discounts, promotions, or data quality issues. Could not incorporate into current model due to uncertainty.

Future Research Opportunity

Investigating price variation patterns could unlock better discount-aware recommendations and promotional strategy optimization.

Thoughtful Design Choices

Implicit Feedback Over Ratings

Spending & quantity ratios reveal true preferences better than explicit ratings.

Two-Level Architecture

Separates efficient candidate generation from business-logic re-ranking.

Handling Zero-Price Items

Kept promotional items as valuable co-purchase signals.

Price Sensitivity Tuning

Soft penalties (not hard filters) maintain diversity while respecting budgets.

Repurchase Cycle Optimization

Boosts items due for repurchase; avoids recently bought items.

Opportunities for Enhancement

With Product Descriptions

Use NLP for content similarity & semantic understanding to improve cold-start.

Enhanced Cold-Start

Leverage browsing history, demographics, and category preferences.

System Scalability

Move to online learning for real-time updates & distributed ALS.

Price Variance Investigation

Analyze extreme fluctuations to distinguish promotions from data errors.

System Scalability

Move to online learning for real-time updates & distributed ALS.