Milestone 3 Report

Course: CAP5771 - Spring 2025

Name: Rushang Sunil Chiplunkar

Date: April 20, 2025

Model Evaluation and Interpretation

Evaluation Methodology

The performance of the different recommendation strategies was evaluated using a methodology implemented within the project's modeling script (`6_model.py`). The core steps involved:

1. **Test Set Simulation:** A random sample of 100 restaurants was drawn from the final dataset (`final_df.csv`) to represent diverse test cases.
2. **Recommendation Generation:** For each restaurant in the sample, the top 10 recommendations were generated using the primary methods developed: Content-Based Filtering, a Personalized approach (using the target restaurant as 'liked'), the location-based Collaborative Filtering (CF) model using SVD, and the Hybrid model.
3. **Metric Calculation:** Two custom metrics were employed to assess recommendation relevance relative to the source restaurant in each test case:
    - **Rating Precision:** Calculated as the proportion of recommended restaurants with a rating within ±0.5 points of the source restaurant's rating. This measures the model's ability to recommend items of a similar perceived quality tier.
    - **Cuisine Match Rate:** Calculated as the proportion of recommended restaurants sharing at least one cuisine type with the source restaurant. This measures topical relevance based on cuisine.
4. **Performance Aggregation:** The individual metric scores were averaged across all test scenarios to provide a summary of each model's typical performance.

Evaluation Metrics Results

The evaluation yielded the following average performance scores for each method, based on the results reported in the project documentation:

Recommendation Method	Avg. Rating Precision	Avg. Cuisine Match Rate
Content-Based	0.8250	0.6400
Personalized	0.8762	0.8713
Hybrid	**0.9180**	**0.8930**
Collaborative Filtering (CF)	0.9980	0.1330

Observations:

- The **Collaborative Filtering** model (based on location-SVD) achieved near-perfect Rating Precision, indicating it strongly recommends restaurants with similar rating levels based on latent location factors. However, its very low Cuisine Match Rate suggests a failure to align recommendations with specific culinary types relevant to the source restaurant.
- The **Content-Based** method showed moderate performance on both metrics.
- The **Personalized** approach significantly improved upon the basic Content-Based method in both precision and cuisine matching.
- The **Hybrid model demonstrated the most balanced performance**, achieving high scores for both Rating Precision (0.9180) and Cuisine Match Rate (0.8930). This indicates its ability to recommend restaurants that are both high-quality (similar rating) and relevant (matching cuisine).
- Based on these results, the **Hybrid model was identified as the preferred approach**, offering the best trade-off between recommending relevant and high-quality options. The optimal weights determined for this model are: Content-Based: 0.60, Collaborative Filtering: 0.10, Popularity-Based: 0.30.

Interpretation and Insights

Analysis of the model components provides the following insights:

1. **Popularity Component:** The engineered `popularity_score`(derived from normalized rating, rating count, and inverse cost) proved to be a dominant feature in predicting restaurant

ratings, as evidenced by feature importance analysis (`7_feature_evaluation.py`). This confirms that established popularity (high ratings, many reviews) is a powerful signal in the dataset. Models leveraging this score effectively capture general appeal.

2. **Content-Based Component:** This component, using TF-IDF on `cuisines` and `location`, excels at identifying restaurants with similar explicit characteristics in comparable areas. It provides explainable recommendations based on textual similarity ("similar cuisine in this area"). Its effectiveness hinges on the descriptive quality of the input text features.
3. **Collaborative Filtering Component (Location-SVD):** The SVD model applied to the location-restaurant matrix identifies latent similarities between locations based on shared rating patterns. While precise in matching rating levels within these latent dimensions, its weak connection to explicit features like cuisine limits its practical relevance when used alone, as shown by the evaluation metrics. It primarily captures geographic taste trends.
4. **Hybrid Model Component:** By combining the different signals, the hybrid model leverages the strengths of each approach. The high weight assigned to Content (0.6) and significant weight to Popularity (0.3) compared to CF (0.1) in the optimal configuration reflects the evaluation findings – content similarity and general popularity were stronger indicators of relevant, quality recommendations than the location-based CF in this specific implementation and dataset. This allows the model to balance finding similar options with suggesting generally well-regarded ones.

Biases and Limitations

Acknowledging the limitations and potential biases of the models is essential for responsible future development:

1. **Data Biases:**
   - *Source Bias:* Reliance on specific platforms (Swiggy, Zomato) introduces their inherent biases regarding listed restaurants and user demographics.
   - *Geographic Bias:* The dataset's focus on Bangalore limits generalizability, and certain areas within the city might be over/under-represented.
   - *Temporal Bias:* The static nature of the dataset means it doesn't reflect real-time changes in ratings, costs, or restaurant availability.
2. **Algorithmic Biases:**
   - *Popularity Bias:* The significant weight of popularity metrics can lead to recommending already famous restaurants, potentially creating filter bubbles and disadvantaging new or niche establishments (item cold-start).
   - *CF Limitations:* Using location as a user proxy is a major simplification, ignoring individual taste variations. The model struggles with new locations (location cold-start) and can be affected by data sparsity.

3. **Feature Limitations:**
   - *Content Quality:* The effectiveness of content filtering depends on the accuracy and detail of cuisines and location data. TF-IDF captures keyword co-occurrence but not deeper semantic meaning.
   - *Missing Features:* The model doesn't incorporate user review text, images, menu details, or real-time signals, which could significantly enhance recommendation quality and context.
4. **Evaluation Limitations:**
   - *Metrics Scope:* Rating Precision and Cuisine Match Rate measure specific relevance aspects but don't capture user satisfaction comprehensively (e.g., diversity, novelty).
   - *Methodology:* Evaluation based on random item sampling doesn't fully replicate real user interaction scenarios or test personalization effectively without actual user feedback data. No true train/test split with user interactions was feasible.
5. **Model Scope:** The hybrid weights, while optimized based on the evaluation, are static and may not be ideal for every user or query context.

Future work could focus on mitigating these limitations through techniques like incorporating user reviews via NLP, exploring methods for bias reduction (e.g., re-ranking), obtaining more diverse or real-time data, and designing more comprehensive online or offline evaluation frameworks if user interaction data becomes available.