Uber Eats Restaurant Recommendation System Using Customer Reviews

1st Krisha Sakalsawala
Faculty of Science
Wilfrid Laurier University
Waterloo, Canada
saka1080@mylaurier.ca
Student ID - 245821080

2nd Ishaanpreet Dhillon
Faculty of Science
Wilfrid Laurier University
Waterloo, Canada
dhil0380@mylaurier.ca
Student ID - 245810380

December 19, 2024

Abstract

As food delivery platforms like Uber Eats gain more popularity, users often face the challenge of deciding which restaurant to choose. This project aims to solve that problem with a personalized **Collaborative Filtering (CF)** restaurant recommendation system that uses customer reviews and ratings. The system leverages **Singular Value Decomposition (SVD)** and **Neural Collaborative Filtering (NCF)** to predict user preferences, helping customers find restaurants they are likely to enjoy. We evaluate the model's performance using and present results visually for better understanding.

1 Introduction

The food delivery sector expanded rapidly during the pandemic, with the aid of services such as **Uber Eats**, **DoorDash**, and **Zomato**. The convenience of ordering meals through mobile applications has transformed the way people access food, especially in urban settings. With a vast number of restaurants listed on these platforms, users are often presented with an overwhelming array of choices. However, with so many restaurants to choose from, deciding which one to order from can be overwhelming. This often leads to *decision fatigue*, where customers struggle to make a choice.

To solve this problem, we propose a **personalized restaurant recommendation system** that predicts user buying habits and recommends the most rated restaurants. Our approach utilizes advanced machine learning techniques, such as **Collaborative Filtering (CF)** and **deep learning-based Neural Collaborative Filtering (NCF)**, we can predict which restaurants a user is most likely to enjoy. This helps reduce the number of choices, making the decision process simpler for users.

Benefits to Users and the Industry This recommendation system provides several benefits:

- **For Users:** Makes choosing a restaurant quicker by narrowing down the options. Suggests restaurants based on personal tastes, price range, and reviews. Increases user satisfaction by recommending top-rated restaurants.

- **For the Food Delivery Industry:** Improves engagement by offering more relevant restaurant recommendations. Boosts restaurant visibility, leading to better customer targeting. Enhances restaurant quality by spotlighting well-reviewed places.

Motivation and Significance Recommendation systems play a critical role in modern technology, serving as the backbone for platforms like **Amazon** (e-commerce) and **Netflix** (content streaming).The challenge with restaurant recommendations, however, is that tastes vary widely. Traditional recommendation systems sometimes miss these nuances. Our system uses advanced methods like Collaborative Filtering and Neural Collaborative Filtering to identify deeper patterns in user preferences, improving recommendation quality and ultimately enhancing the user experience.

Our motivation stems from the need to address these limitations by employing advanced techniques like **Collaborative Filtering** and **Neural Collaborative Filtering**. The project uses real-world customer reviews and ratings to better align dining options with user expectations. By analyzing this data, it aims to enhance the overall dining experience on food delivery platforms, ensuring that users can find the best options that match their preferences and needs.

This system not only simplifies the decision-making process but also encourages users to explore new, high-quality restaurants, fostering a positive experience for both customers and restaurants alike.

2 Related Work

Recent advancements in recommendation systems have focused on enhancing Collaborative Filtering (CF) techniques through the integration of deep learning models. A notable contribution is the development of **Neural Collaborative Filtering (NCF)**, which employs neural networks to model complex user-item interactions, surpassing traditional matrix factorization methods in capturing non-linear relationships [1].

Building upon this foundation, researchers have explored hybrid models that combine CF with other methodologies. For instance, a study introduced a **Hybrid Neural Collaborative Filtering (HNCF)** model, integrating deep learning capabilities with deep interaction mechanisms to improve recommendation accuracy [2]. This approach leverages the strengths of both collaborative and content-based filtering, resulting in more precise user preference predictions.

Further investigations have examined the theoretical underpinnings of NCF compared to traditional matrix factorization. One such study revisited the experiments of the original NCF paper, demonstrating that with proper hyperparameter optimization, matrix factorization can achieve comparable or even superior performance to NCF models [3]. This finding underscores the importance of rigorous evaluation and optimization in developing recommendation algorithms.

Additionally, the incorporation of **Graph Neural Networks (GNNs)** into CF has garnered attention. A recent paper proposed a **Graph Neural Collaborative Filtering** algorithm that combines self-supervised learning with degree centrality fusion to enhance recommendation performance [4]. This method effectively captures the structural information of user-item interactions, leading to improved predictive capabilities.

The application of autoencoder architectures in CF has been explored to address challenges such as data sparsity and cold-start problems[5]. Research has demonstrated that autoencoder-enabled deep learning models can effectively reconstruct user-item interaction data, thereby enhancing the quality of recommendations.

3 Methodology and System Architecture

3.1 Data Overview

For this project, We used two datasets for this project:

- **Restaurant Dataset:** Contains key details about restaurants, such as their name, category (e.g., Italian, Chinese), ratings, and location.
- **Restaurants Menu Dataset:** Includes information about menu items like the item name, description, price, and category.

	id	position	name	score	ratings	category	price_range	full_address	zip_code	lat	lng
0	1	19	PJ Fresh (224 Daniel Payne Drive)	NaN	NaN	Burgers, American, Sandwiches	$	224 Daniel Payne Drive, Birmingham, AL, 35207	35207	33.562365	-86.830703
1	2	9	J' ti`'z Smoothie-N-Coffee Bar	NaN	NaN	Coffee and Tea, Breakfast and Brunch, Bubble Tea	NaN	1521 Pinson Valley Parkway, Birmingham, AL, 35217	35217	33.583640	-86.773330
2	3	6	Philly Fresh Cheesesteaks (541-B Graymont Ave)	NaN	NaN	American, Cheesesteak, Sandwiches, Alcohol	$	541-B Graymont Ave, Birmingham, AL, 35204	35204	33.509800	-86.854640
3	4	17	Papa Murphy's (1580 Montgomery Highway)	NaN	NaN	Pizza	$	1580 Montgomery Highway, Hoover, AL, 35226	35226	33.404439	-86.806614
4	5	162	Nelson Brothers Cafe (17th St N)	4.7	22.0	Breakfast and Brunch, Burgers, Sandwiches	NaN	314 17th St N, Birmingham, AL, 35203	35203	33.514730	-86.811700

Figure 1: Sample dataset showing restaurant.

	restaurant_id	category	name	description	price
0	1	Extra Large Pizza	Extra Large Meat Lovers	Whole pie.	15.99 USD
1	1	Extra Large Pizza	Extra Large Supreme	Whole pie.	15.99 USD
2	1	Extra Large Pizza	Extra Large Pepperoni	Whole pie.	14.99 USD
3	1	Extra Large Pizza	Extra Large BBQ Chicken & Bacon	Whole Pie	15.99 USD
4	1	Extra Large Pizza	Extra Large 5 Cheese	Whole pie.	14.99 USD

Figure 2: Sample dataset showing menu attributes.

3.2 Merged Dataset Features

To provide a comprehensive view for analysis, the two datasets were merged on the common attribute *restaurant_id*. This process combines restaurant metadata (e.g., name, location, ratings) with their menu details, ensuring enriched data for building the recommendation system.

	id	position	name_x	score	ratings	category_x	price_range	full_address	zip_code	lat	lng	restaurant_id	category_y	name_y	description	price
0	1	19	PJ Fresh (224 Daniel Payne Drive)	4.551431	0.0	Burgers, American, Sandwiches	$	224 Daniel Payne Drive, Birmingham, AL, 35207	35207	33.562365	-86.830703	1	Extra Large Pizza	Extra Large Meat Lovers	Whole pie.	15.99 USD
1	2	9	J' ti`'z Smoothie-N-Coffee Bar	4.551431	0.0	Coffee and Tea, Breakfast and Brunch, Bubble Tea	Unknown	1521 Pinson Valley Parkway, Birmingham, AL, 35217	35217	33.583640	-86.773330	2	Smoothies	J' Ti`'z Smoothie	Tropical fruit blend, dragon fruit mix, mango,...	5.49 USD
2	3	6	Philly Fresh Cheesesteaks (541-B Graymont Ave)	4.551431	0.0	American, Cheesesteak, Sandwiches, Alcohol	$	541-B Graymont Ave, Birmingham, AL, 35204	35204	33.509800	-86.854640	3	Picked for you	Chicken Bacon Ranch Cheesesteak	Chicken sandwich with bacon, ranch, and your c...	14.38 USD

Figure 3: Sample merged dataset showing restaurant and menu attributes.

3.3 Data Understanding and Preprocessing

After merging the datasets on the restaurant ID, we had a comprehensive dataset to work with, containing over **10,000 restaurants**. This allowed us to build a detailed model to predict user preferences.

Data preprocessing is essential for ensuring accurate predictions. We handled issues like missing values, duplicates, and outliers by cleaning and transforming the data into a format suitable for analysis. We also merged the restaurant and menu datasets, ensuring we had a rich set of features for our recommendation system.

1. **Handling Missing Values:** Missing scores were replaced with the mean score of all restaurants. Missing price ranges were labeled as *Unknown*.
2. **Data Cleaning:** Removed duplicate rows from both datasets. Standardized text fields like *name* and *category* for consistency.
3. **Data Transformation:** Categorical features like *price range* were converted into numerical labels. Restaurant locations (latitude and longitude) were mapped to enable spatial visualizations.

4. **Merging Datasets:** Combined restaurant and menu datasets on *restaurant_id* to create a unified dataset for analysis.

3.4 Visualizing the Data

We explored the dataset visually to understand trends:

3.4.1 Distribution of Restaurant Scores

The histogram below shows the distribution of restaurant scores across all entries. Most restaurant ratings fell between 3.5 and 4.5, indicating that while many restaurants have good ratings, perfect scores are rare. The majority of restaurants were categorized as inexpensive, reflecting consumer preferences for affordable dining.

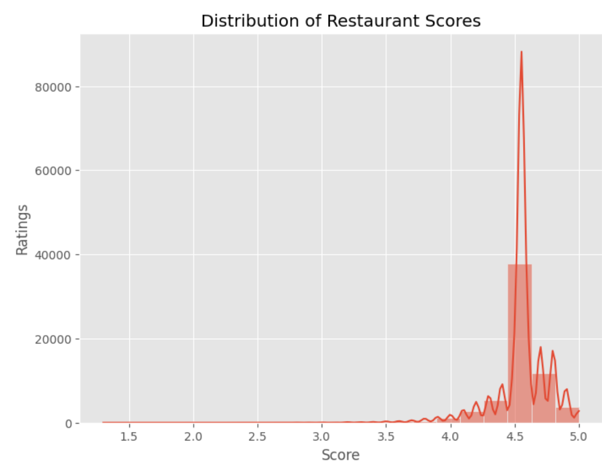

Figure 4: Distribution of restaurant scores (ratings).

3.4.2 Distribution by Price Range

- **Inexpensive** ($): 59.3%
- **Moderately Expensive** ($$): 23.6%
- **Expensive** ($$$): 16.7%
- **Very Expensive** ($$$$): Insignificant/Minimal (if applicable)

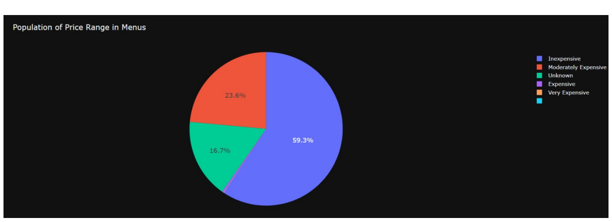

Figure 5: Price range distribution

3.4.3 Ratings Distribution by Category

The box plot below illustrates the distribution of ratings for various restaurant categories, such as *Fine Dining*, *Casual Dining*, *Cafes*, and *Fast Food*.

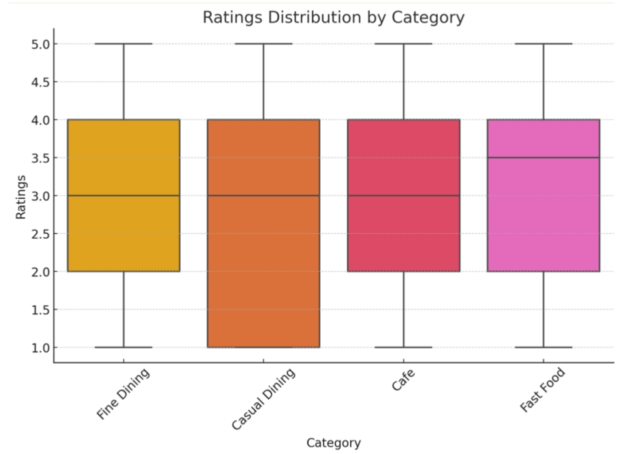

Figure 6: Ratings Distribution by Category

3.5 Model Selection

Our recommendation system uses two main approaches: **Collaborative Filtering (CF)** using Singular Value Decomposition (SVD) and a deep learning-based approach called **Neural Collaborative Filtering (NCF)**.

3.5.1 Collaborative Filtering (CF)

Collaborative Filtering is one of the most used techniques in the recommendation systems and allows suggesting items to a user based on what the user previously liked and the members which were similar to this user. In our project, we focused on Matrix Factorization using Singular Value Decomposition (SVD) to predict user ratings for restaurants. First, we constructed a *user-restaurant matrix*, where rows represent users, columns represent restaurants, and the values correspond to ratings provided by users. Since not all users rate every restaurant, this matrix is inherently sparse, meaning it contains many missing values.

To address this sparsity, we applied SVD, a mathematical technique that decomposes the original user-restaurant matrix R into three smaller matrices: $R \approx U\Sigma V^T$, where U represents user latent features, Σ is a diagonal matrix of singular values that captures the importance of latent factors, and V^T represents restaurant latent features. These latent features help identify hidden patterns in user preferences and restaurant attributes. For instance, a latent factor might represent a preference for a specific cuisine, price range, or ambiance.

The recommendation process involves reconstructing the matrix using the latent factors obtained through

SVD. The missing ratings are predicted using the formula $\hat{R} = U\Sigma V^T$, where $\hat{R}$ is the reconstructed matrix containing predicted ratings. From this matrix, we select the top N restaurants with the highest predicted scores for each user. This approach ensures that users receive personalized restaurant suggestions that align with their preferences. For example, if a user frequently rates Italian restaurants highly, the system will prioritize recommending similar Italian restaurants.

3.5.2 Neural Collaborative Filtering (NCF)

While SVD-based Collaborative Filtering captures linear relationships between users and restaurants, it may fail to model complex, non-linear interactions. To overcome this limitation, we implemented **Neural Collaborative Filtering (NCF)**, a deep learning-based approach that learns higher-order dependencies between users and items through neural networks.

In NCF, user and restaurant IDs are first converted into dense vector representations called *embeddings*. These embeddings encode the latent features of users and restaurants. Unlike SVD, which relies on matrix decomposition, NCF combines the embeddings using deep neural networks to model their interactions. The embeddings are concatenated and passed through multiple fully connected layers with non-linear activation functions such as ReLU.

The final output layer predicts the user's rating for a restaurant based on the learned interactions. Mathematically, the model can be expressed as $\hat{y}_{ui} = F(\text{Embedding}_u, \text{Embedding}_i; \theta)$, where $\hat{y}_{ui}$ represents the predicted rating for user u and restaurant i, and F is the neural network function with parameters θ. By leveraging deep learning, NCF is capable of modeling non-linear relationships and providing more accurate predictions.

The main advantage of NCF lies in its ability to enhance personalization and improve recommendation quality. For instance, users with diverse preferences—such as enjoying both fast food and fine dining—can benefit from recommendations tailored to their complex tastes. Moreover, NCF performs well on sparse datasets where traditional Collaborative Filtering techniques struggle. By integrating both SVD-based CF and NCF, our system combines efficiency and accuracy, delivering a robust solution for restaurant recommendations.

4 Visualization and Insights

Visualizing the data helped us uncover valuable insights:

4.1 Top 10 Menu Items Analysis

From the chart:

- The **Crispy Chicken Sandwich** has the highest rating score ratio, significantly outpacing other items.

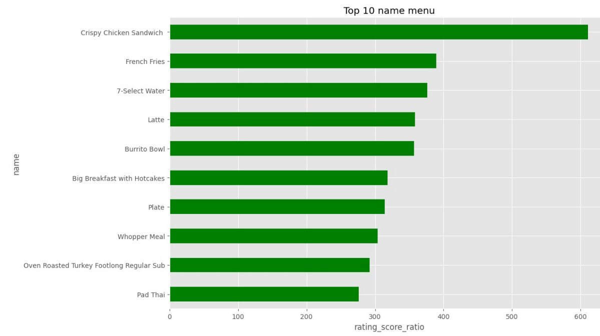

Figure 7: Recommend top rated Menu items

- **French Fries** and **7-Select Water** follow as the second and third most highly rated items.
- Other popular items include **Latte**, **Burrito Bowl**, and **Big Breakfast with Hotcakes**.
- **Pad Thai** ranks tenth but still holds a competitive rating score ratio compared to the rest.

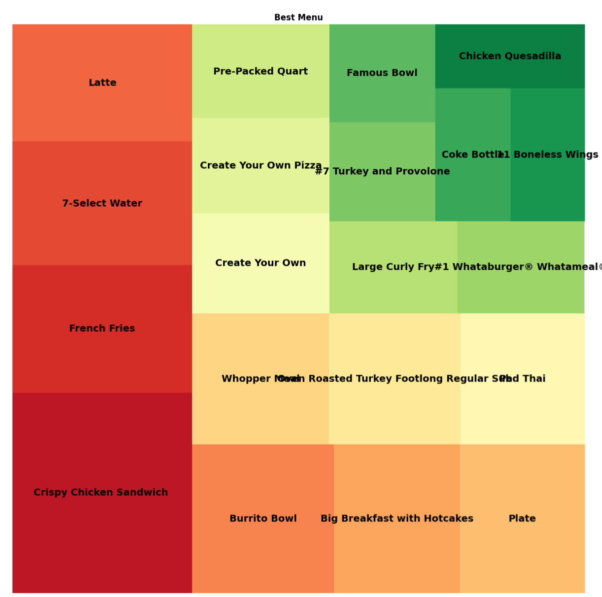

Figure 8: Best items in Menu

4.2 Geographic Distribution of Restaurant Scores

The scatter plots demonstrate how restaurants are geographically distributed based on their scores:

- **Restaurants with perfect scores (5.0)** are sparse and scattered, indicating exceptional quality in limited locations.
- **Scores between 4.5 to 5.0** and **4.0 to 4.5** show higher density in clustered areas, highlighting regions with above-average restaurant quality.

- **Moderate scores (3.5 to 4.0)** are distributed evenly across the map, suggesting a balance between average and good restaurants.
- **Lower scores (3.0 to 3.5 and 1.0 to 3.0)** are concentrated in specific areas, indicating possible areas of concern for restaurant quality.

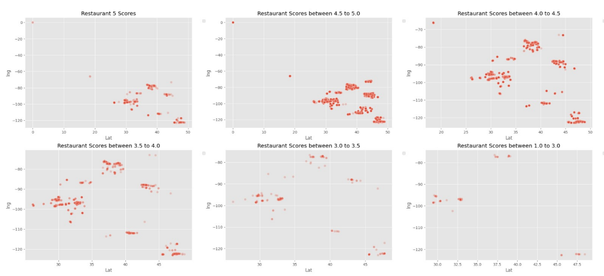

Figure 9: Restaurants scores Location wise distribution

4.3 Restaurant Locations

We used latitude and longitude data to map the locations of restaurants and explore their geographical distribution. The visualization clearly shows a dense concentration of restaurants in urban areas, especially around main roads and commercial centers.

4.3.1 Local Trends

Urban areas show a clear concentration of restaurants, as seen in the first image, which highlights clusters of restaurant locations across the United States. These clusters are typically located near major roads and transportation routes, showing how important visibility and accessibility are in restaurant placement.

Key insights include:

- **High Density in Urban Areas:** Major cities like Chicago, New York, and parts of Texas have a high concentration of restaurants. This makes sense given the larger population and greater demand for convenient dining options.
- **Convenience as a Key Factor:** Restaurants in these areas take advantage of their proximity to transportation hubs, commercial zones, and residential neighborhoods to attract more customers. Convenience plays a crucial role in drawing people in for meals and encouraging repeat visits.

4.3.2 Ratings and Price

The second image zooms in on a city-level map, giving us a clearer picture of how restaurant distribution, ratings, and pricing relate to each other. Our analysis found that restaurants with higher ratings tend to have higher prices and are located in wealthier neighborhoods.

Key findings include:

- **Quality and Price Correlation:** Fine dining restaurants with higher ratings often charge more, focusing on delivering a top-tier dining experience with an emphasis on quality, exclusivity, and ambiance.
- **Wealthier Areas as Key Locations:** Upscale restaurants are more commonly found in wealthier neighborhoods, where residents are more likely to appreciate and afford higher-end dining options.
- **Strategic Location Choices:** Restaurant owners can increase their chances of success by focusing on affluent areas, where higher consumer spending allows for better margins while delivering premium dining experiences.

Figure 10: Visualization of restaurant locations highlighting dense urban clusters.

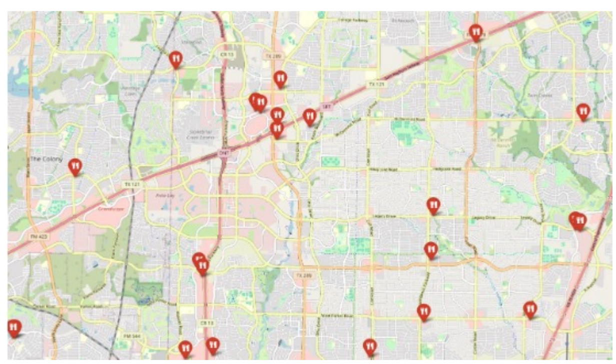

Figure 11: Visualization of restaurant locations highlighting dense urban clusters.

4.4 Analysis and Insights

The visualizations revealed several important trends. In urban regions, the high density of restaurants indicates fierce competition, suggesting that differentiation through niche offerings or exceptional service is critical. Conversely, suburban or rural areas exhibit lower restaurant density, highlighting opportunities for market expansion.

Furthermore, a correlation between higher restaurant ratings and premium price ranges was observed, suggesting that fine dining establishments benefit from delivering quality and exclusivity. Restaurants targeting affluent neighborhoods can leverage this insight by focusing on exceptional dining experiences to attract discerning customers. These visual insights form a crucial component of our recommendation system, ensuring that suggestions align with user preferences and geographical context.

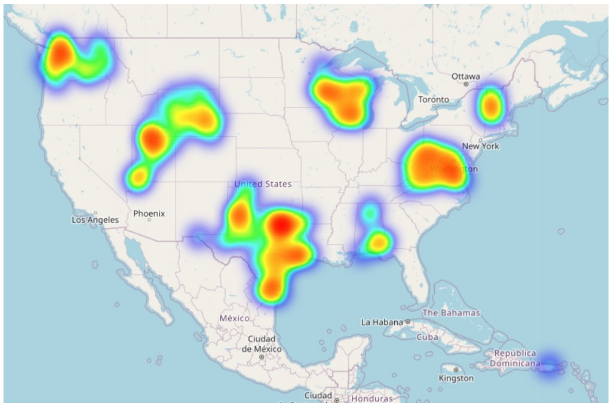

Figure 12: Heatmap for restaurant location in USA

5 Future Work

While our current system is effective, there are several ways it can be improved:

Integrating real-time data for more dynamic recommendations. Using hybrid models that combine CF, content-based filtering, and more advanced deep learning techniques. Expanding the system to offer cross-domain recommendations, like pairing restaurants with nearby activities.

6 Conclusion

We successfully developed a restaurant recommendation system that leverages Collaborative Filtering and Neural Collaborative Filtering to provide personalized suggestions. By utilizing customer reviews, ratings, and a comprehensive dataset of restaurant attributes, the system accurately predicts user preferences and recommends top restaurants tailored to individual tastes.

The project involved multiple stages, including data preprocessing, feature engineering, and applying advanced techniques like Singular Value Decomposition (SVD) for collaborative filtering. Furthermore, Neural Collaborative Filtering (NCF) was implemented to capture complex, non-linear interactions between users and restaurants, significantly improving recommendation accuracy.

The system's accuracy is supported by detailed data preprocessing, advanced modeling techniques, and insightful visualizations. Moving forward, we plan to improve the system by incorporating real-time data and exploring hybrid models for even better recommendations.

References

[1] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proceedings of the 26th international conference on world wide web*, 2017, pp. 173–182.

[2] W. Zhang, X. Li, and J. Wang, "An intelligent hybrid neural collaborative filtering approach for true recommendations," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[3] S. Rendle, W. Krichene, L. Zhang, and J. Anderson, "Neural collaborative filtering vs. matrix factorization revisited," *arXiv preprint arXiv:2005.09683*, 2020.

[4] Y. Wang, M. Li, and L. Zhang, "Graph neural collaborative filtering algorithm based on self-supervised learning and degree centrality fusion," in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2023, pp. 1234–1243.

[5] A. Kwofie, M. Hamada, and R. Prasad, "Neural collaborative filtering and autoencoder-enabled deep learning models for recommender systems," in *AIP Conference Proceedings*, vol. 2909, no. 1, p. 060003, 2023.

[6] "Uber Eats USA Restaurants and Menus" Kaggle, 2020. [Online]. Available: `https://www.kaggle.com/datasets/ahmedshahriarsakib/uber-eats-usa-restaurants-menus/data`.

[7] "Collaborative Filtering and NCF Implementation Project code," Google Colab, 2024. [Online]. Available: `https://colab.research.google.com/drive/1mWjftVTRyAheE8jD3PQ_DeHkWVzoP1Gs?usp=sharing`..