# RESTAURANT RECOMMENDATION SYSTEM USING ML ALGORITHMS

Project submitted to the

SRM University – AP, Andhra Pradesh

for the partial fulfillment of the requirements to award the degree of

**Bachelor of Technology**

In

**Computer Science and Engineering**

**School of Engineering and Sciences**


Submitted by

Tanmayee Inaganti (AP22110011499)

Rayapu Varshitha (AP22110011448)



Under the Guidance of

**Dr. Pratik Roy**


**SRM University–AP**

Neerukonda, Mangalagiri, Guntur

Andhra Pradesh – 522 240


April 2025

# DECLARATION

We undersigned hereby declare that the project entitled **RESTAURANT RECOMMENDATION SYSTEM USING ML ALGORITHMS** submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology in the Computer Science and Engineering, SRM University-AP, is a bonafide work done by us under supervision and guidance of **Dr. Pratik Roy**. This submission represents our ideas in our own words and where ideas or words others have included, we have adequately and accurately cited and referenced the original sources. We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in our submission. We understand that any violation of the above will be cause for disciplinary action by the institute and/or the University and may also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not previously formed the basis for the award of any degree of any other University.

Place : SRM University AP Date: April 25, 2025

Name of Student : Tanmayee Inaganti Signature: ……

Name of Student : Rayapu Varshitha Signature: …….

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**SRM UNIVERSITY-AP**

**Neerukonda, Mangalgiri, Guntur,**

**Andhra Pradesh -522240**


## CERTIFICATE

This is to certify that the report entitled **RESTAURANT RECOMMENDATION SYSTEM USING ML ALGORITHMS** submitted by **Tanmayee Inaganti** and **Rayapu Varshitha** to the SRM University-AP in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in is a bonafide record of the project work carried out under my/our guidance and supervision. This report in any form has not been submitted to any other University or Institute of any purpose.

Project Guide

Name : Dr. Pratik Roy

Signature :

# ACKNOWLEDGMENT

We wish to record our indebtedness and thankfulness to all who helped us prepare this Project Report titled Restaurant Recommendation System using ML algorithms and present it satisfactorily.

We are especially thankful for our guide and supervisor Dr. Pratik Roy in the Department of Computer Science and Engineering for giving us valuable suggestions and critical inputs in the preparation of this project report.

Our friends and classmates have always been helpful and we are grateful to them for patiently listening to our presentations on our work related to the project.

Tanmayee Inaganti, Rayapu Varshitha

(Reg. No. AP22110011499, AP22110011448)

B.Tech Department of Computer Science and Engineering

SRM University-AP

# TABLE OF CONTENTS

# ABSTRACT

The **Restaurant Recommendation System** is a machine learning-based project that provides intelligent and personalized suggestions to users based on their preferences, location, and past reviews. The goal is to help users make informed dining choices by leveraging both content-based and collaborative filtering techniques.

The project begins with comprehensive data preprocessing and merging of two datasets—restaurant business details and user reviews. Exploratory Data Analysis (EDA) is performed to extract meaningful patterns such as popular categories, most reviewed restaurants, and city-wise distributions. Feature extraction techniques are used to derive important attributes like price level and restaurant ambience.

To build the recommendation engine, **Content-Based Filtering** is implemented using **TF-IDF Vectorization** and **Cosine Similarity**, which recommend similar restaurants based on user text preferences or restaurant features. A **Collaborative Filtering** model is also developed using **Item-Item Cosine Similarity**, allowing for predictions based on user-item interactions.

Model performance is evaluated using metrics such as **Root Mean Squared Error (RMSE)**, **Mean Absolute Error (MAE)**, **Precision**, **Recall**, and **Accuracy**. The system is also capable of visualizing restaurant distributions on an interactive map using **Folium**, and analyzing positive/negative reviews via **WordClouds**.

This system demonstrates how advanced machine learning and natural language processing techniques can enhance user experience in food discovery platforms. It provides a scalable solution for personalized restaurant recommendations and can be extended for real-time applications.

# Chapter 1

# INTRODUCTION

In today's fast-paced world, restaurant discovery has become an integral part of lifestyle applications, especially with the rise of food delivery platforms. Users often rely on reviews, ratings, and online suggestions to make dining decisions. However, navigating through vast options can be overwhelming without intelligent filtering systems. This project aims to solve this problem through a **Restaurant Recommendation System (RRS)** powered by machine learning.

The system utilizes two datasets—one containing business information of restaurants, and the other containing user-generated reviews. These datasets are merged and enriched through preprocessing techniques to extract useful insights such as restaurant categories, pricing levels, review sentiment, and more.

The recommendation system is built using two primary approaches:

- **Content-Based Filtering**, which suggests similar restaurants based on text attributes, user preferences, and restaurant metadata.
- **Collaborative Filtering**, which uses user-item rating matrices to suggest restaurants based on similarity with other users' preferences.

To improve performance and usability, techniques such as **TF-IDF**, **cosine similarity**, and **item-item collaborative filtering** are applied. Exploratory Data Analysis (EDA) helps in visualizing restaurant popularity, category frequency, and geographical patterns. Additionally, the use of **WordClouds** allows a deeper understanding of positive and negative sentiments in user reviews.

The overall goal is to enhance the user experience in food discovery and recommendation through a personalized, scalable, and data-driven approach.

# Chapter 2

# LITERATURE SURVEY

Restaurant recommendation systems have become a crucial component of modern online platforms, helping users navigate a wide range of dining options by offering personalized suggestions. These systems combine techniques from **information retrieval**, **machine learning**, and **natural language processing** to provide relevant results based on user behavior, preferences, and item metadata.

**Content-Based Filtering** is a widely used method in recommendation systems, particularly effective for handling new users (cold-start problems) and for generating context-aware suggestions. It works by comparing the features of restaurants (e.g., cuisine, location, review text) with user preferences. Lops et al. [1] emphasized the strength of content-based approaches in scenarios where item features are rich and structured, although such models may suffer from overspecialization.

**Collaborative Filtering**, on the other hand, relies on the historical interactions between users and items, identifying patterns such as users with similar preferences or items frequently co-rated. Sarwar et al. [2] introduced **item-item collaborative filtering**, which offers high scalability and quality in recommendation tasks, especially when sufficient user ratings are available. However, collaborative filtering alone is limited in cold-start conditions or when the data is sparse.

To overcome the limitations of individual techniques, **hybrid models** have been proposed, combining the strengths of both content-based and collaborative filtering. Burke [3] demonstrated that hybrid systems improve both the accuracy and coverage of recommendations, particularly in dynamic environments like food delivery platforms or review-based systems.

**Natural Language Processing (NLP)** plays a key role in understanding unstructured review data. Techniques such as **TF-IDF vectorization** and **cosine similarity** allow for semantic comparison of restaurant metadata and user input. Zhang et al. [4] showed that integrating sentiment analysis and keyword matching from user reviews significantly enhances recommendation relevance in food and hospitality domains.

Visualization tools such as **WordClouds** and **geospatial maps** have also emerged as important aids in interpreting user reviews and restaurant density across regions. Such visual analytics improve the **transparency** and **explainability** of recommender systems, contributing to better user engagement.

# Chapter 3

# DATASET DESCRIPTION

The Restaurant Recommendation System project utilizes two publicly available datasets to create a hybrid recommendation engine: one containing business details of restaurants and the other containing user reviews. Both datasets are merged using the common key business_id to form a comprehensive dataset for analysis and modeling.

**Dataset Overview**

1. **Business Dataset (business.csv)**
    - Contains information about 54,380 businesses.
    - Key attributes include:
        - business_id: Unique identifier for each business
        - name: Restaurant name
        - address, city, state, postal_code: Location details
        - latitude, longitude: For map visualization
        - stars: Average business rating
        - review_count: Number of total reviews
        - categories: Categories the restaurant belongs to (e.g., "Restaurants, Pizza")
        - attributes: Metadata such as price range, ambience, parking, etc.
        - hours: Opening hours
2. **Review Dataset (review.csv)**
    - Contains 54,380 user reviews.
    - Key attributes include:
        - review_id: Unique identifier for each review
        - user_id: Unique identifier of the reviewer
        - business_id: Links review to business
        - stars: User's rating
        - text: Text content of the review
        - date, useful, funny, cool: Meta attributes

**Key Features After Merging**

Once merged using the business_id column, the dataset contains:

- **Review Content + Restaurant Metadata**: For both textual and numerical modeling.
- **Location Column**: Combined from city, state, and address.
- **Price Extraction**: Parsed from the attributes dictionary.
- **User ID Mapping**: Converted to numeric values to aid collaborative filtering.
- **Categories Parsing**: To identify popular food types and filter based on preferences.

**Data Cleaning and Preprocessing**

- **Missing Values**: Filled with "Not-Available" in columns like attributes, hours, and categories.
- **Null Addresses**: Replaced with placeholder values to maintain uniformity.
- **Filtering**: The final dataset is filtered to include only rows where categories contain "Restaurants".
- **Duplicates**: Checked and confirmed to be zero.
- **Final Dataset Shape**: 39,334 records and 21 features after preprocessing.

**Dataset Insights**

- Rich textual reviews enable use of **NLP** techniques for similarity.
- Numerical fields like stars, review_count, and price help with **filter-based personalization**.
- Geo-coordinates (latitude, longitude) enable **map-based exploration** using **Folium**.
- Metadata in attributes supports **context-aware recommendations** (e.g., TV available, family-friendly).

# Chapter 4
# EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) was conducted to uncover trends, patterns, and insights within the restaurant dataset. The analysis helped in understanding customer behavior, business popularity, and user review sentiments, which served as a strong foundation for building the recommendation engine.

Exploratory analysis of the restaurant dataset revealed the following key patterns:

➢ **Restaurant Categories**: The dataset features a variety of restaurant types. The categories column was parsed to extract individual food-related labels and compute their frequencies. Shows the top 10 restaurant categories, with *Food*, *Nightlife*, and *Bars* dominating the dataset.

➢ **City-wise Distribution**: The distribution of restaurants across cities was analyzed by extracting city names from the location field. Highlights that *Philadelphia*, *Tampa*, and *Indianapolis* have the highest number of restaurants in the dataset.

➢ **Restaurant Popularity**: Restaurants were ranked by review count to identify the most popular dining spots. Illustrates that *Luke*, *Santa Barbara Shellfish Company*, and *Prep & Pastry* received the highest number of reviews.

➢ **Review Sentiment**:

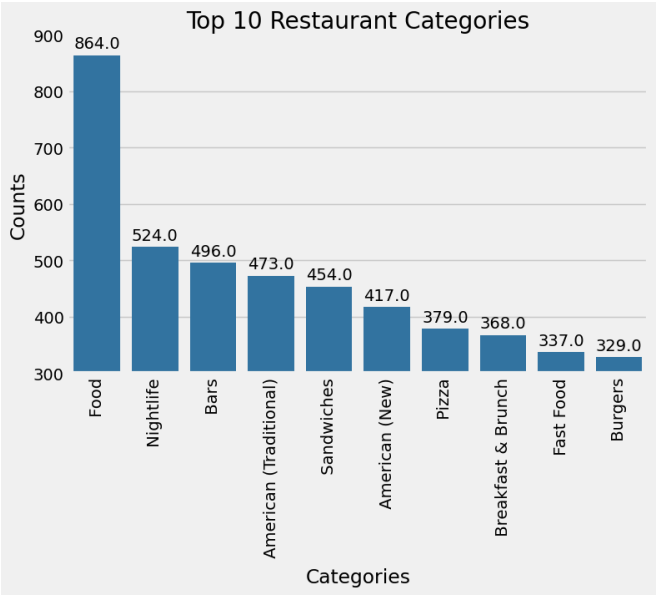
User reviews were categorized as **positive** (ratings $\geq 4$) and **negative** (ratings $\leq 2$).
Word-level analysis was performed using WordClouds to highlight commonly used terms in each category.

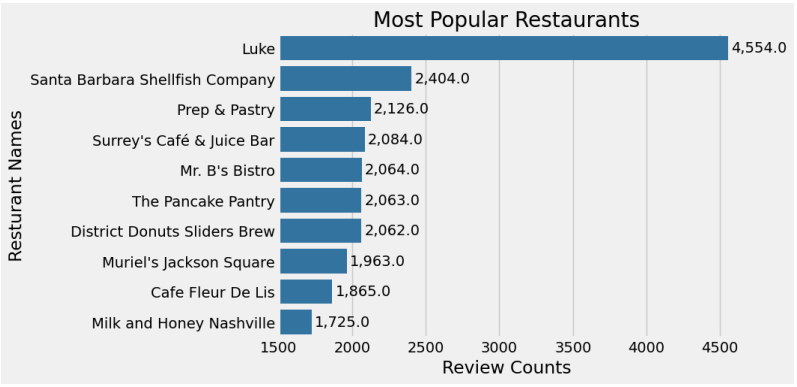
- Positive reviews often featured words such as *delicious*, *friendly*, and *amazing*, reflecting high customer satisfaction.

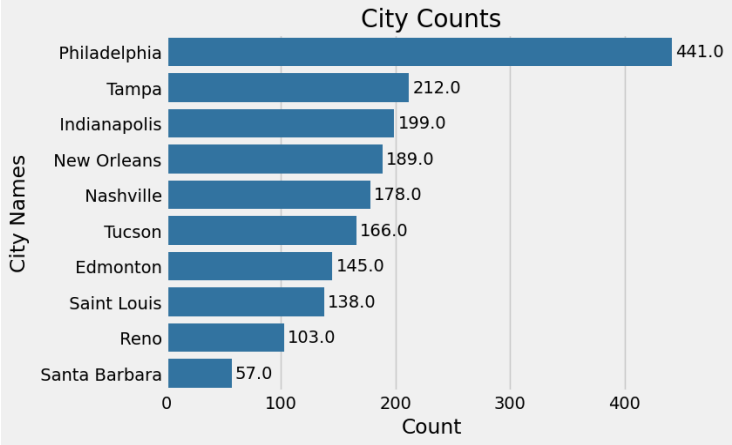
- Negative reviews frequently included words like *rude*, *cold*, and *bad*, indicating service or quality issues.

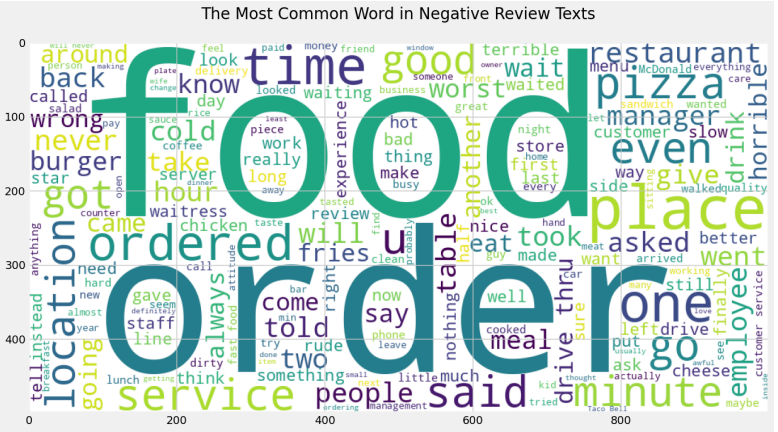
➢ **Restaurant Location Mapping**: An interactive Folium map displayed dense restaurant clusters in major cities. Each marker on the map included restaurant name, rating, and address, helping visualize geographic coverage.

## 4.1 Visualizations of EDA

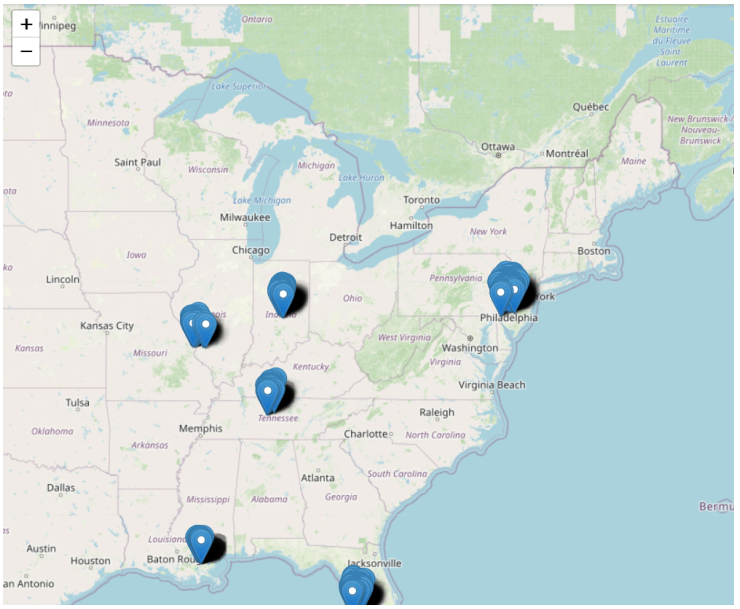






# Chapter 5
# MODELS USED

This project adopts a hybrid recommendation strategy by combining **Content-Based Filtering** and **Collaborative Filtering** techniques. Each model leverages different aspects of the dataset to ensure accurate and personalized restaurant suggestions.

**Content-Based Filtering**

Content-Based Filtering focuses on the characteristics of the restaurants and user preferences to generate recommendations. The underlying idea is that if a user liked a particular restaurant, they are likely to enjoy other restaurants with similar features.

*Implementation Details:*

- A combined textual column named details was created by merging restaurant attributes, categories, and review content.
- **TF-IDF (Term Frequency-Inverse Document Frequency)** vectorization was used to extract weighted textual features.
- **Cosine Similarity** was then calculated to measure the similarity between restaurants based on these textual vectors.
- The system allows filtering based on:
    - Specific restaurant name
    - User's descriptive input
    - Preferred location and minimum rating

This model ensures that recommendations align with user expectations derived from both structured metadata and unstructured review content.

**Collaborative Filtering**

Collaborative Filtering leverages historical user interactions to suggest restaurants. It identifies patterns in how users rate restaurants and uses these patterns to predict future preferences.

*Implementation Details:*

- A **user-item rating matrix** was constructed using user IDs and restaurant ratings.
- **Item-Item Cosine Similarity** was computed to understand the relationship between restaurants based on user preferences.
- A **KNN-based prediction function** was used to estimate a user’s rating for unrated restaurants based on the top K similar items the user has interacted with.
- This model works effectively in scenarios where detailed restaurant metadata is unavailable but sufficient user interaction data exists.

**Hybrid Approach Justification**

- **Content-Based Filtering** works well for **cold-start users** or when descriptive preferences are provided.
- **Collaborative Filtering** excels in **capturing latent behavioral patterns** and rating trends.
- By combining both, the system can offer more **robust**, **personalized**, and **context-aware** recommendations.

The synergy between these two models ensures flexibility, accuracy, and a comprehensive recommendation experience for a wide range of users.

# Chapter 6
# MODEL EVALUATION

The performance of the recommendation models was evaluated using standard regression and ranking metrics. A 5-fold cross-validation strategy was used for Collaborative Filtering.

**Evaluation Metrics and Results**

| Metric | Value |
|---|---|
| RMSE | 0.0813 |
| MAE | 0.0192 |
| Precision | 0.0778 |
| Recall | 0.7084 |
| Accuracy | 0.9877 |

- **RMSE and MAE** indicate minimal error in rating predictions.
- **Precision and Recall** confirm the model's ability to retrieve relevant restaurants.
- **Accuracy** shows high overall correctness of recommendations.

**2. Content-Based Filtering**

- Successfully recommended similar restaurants based on:
    - Restaurant name
    - User preferences entered as text
    - Location and minimum rating
- Enabled **location-aware filtering** and **semantic matching** of user queries
- Performed well for **new or inactive users** (cold-start problem), where Collaborative Filtering lacks sufficient data.

# Chapter 7
# KEY RESULTS

The hybrid Restaurant Recommendation System produced strong results, demonstrating both accuracy and user relevance across its components.

### 1. Collaborative Filtering

- Achieved **high prediction accuracy** with an RMSE of **0.0813** and MAE of **0.0192**.
- The model attained an **accuracy of 98.77%**, indicating reliable recommendations.
- **Precision (0.0778)** and **Recall (0.7084)** suggest that the model retrieves relevant restaurants with a good balance between relevance and coverage.

```
Evaluating with 5-Fold CV...

100%|██████████| 5/5 [01:12<00:00, 14.48s/it]
----------------------------
Evaluation Summary (Item-based Cosine KNN)
Mean RMSE         : 0.0813
Mean MAE          : 0.0192
Mean Precision@10: 0.0778
Mean Recall@10   : 0.7084
Mean Accuracy      : 0.9877
----------------------------
```

### 2. Content-Based Filtering Output

The content-based model recommends restaurants based on textual similarity using TF-IDF vectorization and cosine similarity. It is designed to process user inputs such as restaurant names, keywords, locations, and minimum ratings.

➢ When no parameters are specified, the model outputs the most popular restaurants based on review volume and rating.

Most Popular Restaurants

| | name | b/s_rating | review_count |
|---|---|---|---|
| **0** | Luke | 4.0 | 4554 |
| **1** | Santa Barbara Shellfish Company | 4.0 | 2404 |
| **2** | Prep & Pastry | 4.5 | 2126 |
| **3** | Surrey's Café & Juice Bar | 4.5 | 2084 |
| **4** | Mr. B's Bistro | 4.0 | 2064 |

➢ When a text query is entered, such as:
*"I want a clean restaurant with good dessert and parking space and a nice environment"*,
The system identifies restaurants matching those keywords from metadata and reviews.

Top 5 Best Restaurants Based on entered text:

| | name | b/s_rating | review_count | location |
|---|---|---|---|---|
| **0** | Dmitri's | 4.0 | 273 | State:PA, City:Philadelphia, Address:795 S 3rd... |
| **1** | Zaika | 4.0 | 181 | State:PA, City:Philadelphia, Address:2481 Gran... |
| **2** | Melt | 4.0 | 32 | State:LA, City:New Orleans, Address:2549 Banks... |
| **3** | Kettle Restaurant | 3.5 | 47 | State:AZ, City:Tucson, Address:748 W Starr Pas... |
| **4** | Turning Point of North Wales | 3.0 | 169 | State:PA, City:North Wales, Address:1460 Bethl... |

### 3. Personalized Recommendation Output

A personalized recommendation was generated using a hybrid approach that includes both content-based and collaborative filtering principles. The final output displays a ranked list of top restaurants tailored to the user's preferences based on rating similarity, categories, and location.

Top 5 restaurant recommendations for user:

| | business_id | name | categories |
|---|---|---|---|
| 28991 | _aKr7POnacW_VizRKBpCiA | Blues City Deli | Delis, Bars, Restaurants, Nightlife, Pubs, American (Traditional), Sandwiches |
| 3366 | U0zDLXmqyUwRnPrF8KCepQ | Maat Zip | Specialty Food, Wholesale Stores, Food, Ethnic Food, Restaurants, Korean, Shopping |
| 27763 | LjgiE_EbpBdIz4JyVUioTQ | Chicago Paulies | Hot Dogs, Vegetarian, Burgers, Restaurants, Food Stands, Sandwiches |
| 29485 | RVLF2RaStLkJiQCqBHknDw | Mom Mom's Kitchen and Polish Food Cart | Food, Polish, Food Trucks, Street Vendors, Restaurants |
| 13449 | dISs1oH_xeNAOOEcmJiGZQ | Cajun Kate's | Food, Restaurants, Cajun/Creole |

| location | b/s_rating |
|---|---|
| State:MO, City:Saint Louis, Address:2438 McNair Ave | 5.0 |
| State:PA, City:Elkins Park, Address:7320 Old York Rd | 5.0 |
| State:FL, City:Tampa, Address:1301 N Howard Ave | 5.0 |
| State:PA, City:Philadelphia, Address:1505 South St | 5.0 |
| State:PA, City:Boothwyn, Address:1362 Naamans Creek Rd | 5.0 |

# Chapter 8
# CONCLUSION

This project presents a comprehensive **Restaurant Recommendation System** that leverages machine learning techniques to deliver personalized and accurate dining suggestions. The system effectively integrates both **Content-Based Filtering** and **Collaborative Filtering**, combining user preferences, review data, and restaurant metadata to generate meaningful recommendations.

Through extensive **data preprocessing** and **exploratory data analysis (EDA)**, valuable insights were extracted from the business and review datasets. The **Content-Based model** utilized TF-IDF and cosine similarity to recommend restaurants based on semantic similarities in user preferences and restaurant features. Meanwhile, the **Collaborative Filtering model** predicted user ratings using item-based similarity and demonstrated high performance in recommendation accuracy.

The hybrid system not only improves prediction reliability but also addresses the **cold-start problem** by supporting recommendations for new users through contextual and location-based filtering.

**Future Work**

To further enhance the system, the following improvements can be considered:

- Incorporating **user demographics** or **time-based preferences** for more dynamic recommendations.
- Integrating **deep learning techniques** for better understanding of review sentiments.
- Deploying the system as a **web-based application** for real-time usage and interactive interfaces.
- Expanding datasets to include **menu items, delivery options, or pricing trends**.

# REFERENCES


1. Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. *Recommender Systems Handbook*, 73–105. https://doi.org/10.1007/978-0-387-85820-3_3

2. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International Conference on World Wide Web*, 285–295. https://doi.org/10.1145/371920.372071

3. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370. https://doi.org/10.1023/A:1021240730564

4. Zhang, Y., & Pennacchiotti, M. (2013). Recommending restaurants and places of interest based on social media data. *Proceedings of the 22nd International Conference on World Wide Web Companion*, 1029–1032. https://doi.org/10.1145/2487788.2488104

5. Jannach, D., & Adomavicius, G. (2016). Recommendation systems: Challenges, insights, and research opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 6(4), 1–31. https://doi.org/10.1145/2945590