# GOURMET GUIDE

# CONTENT



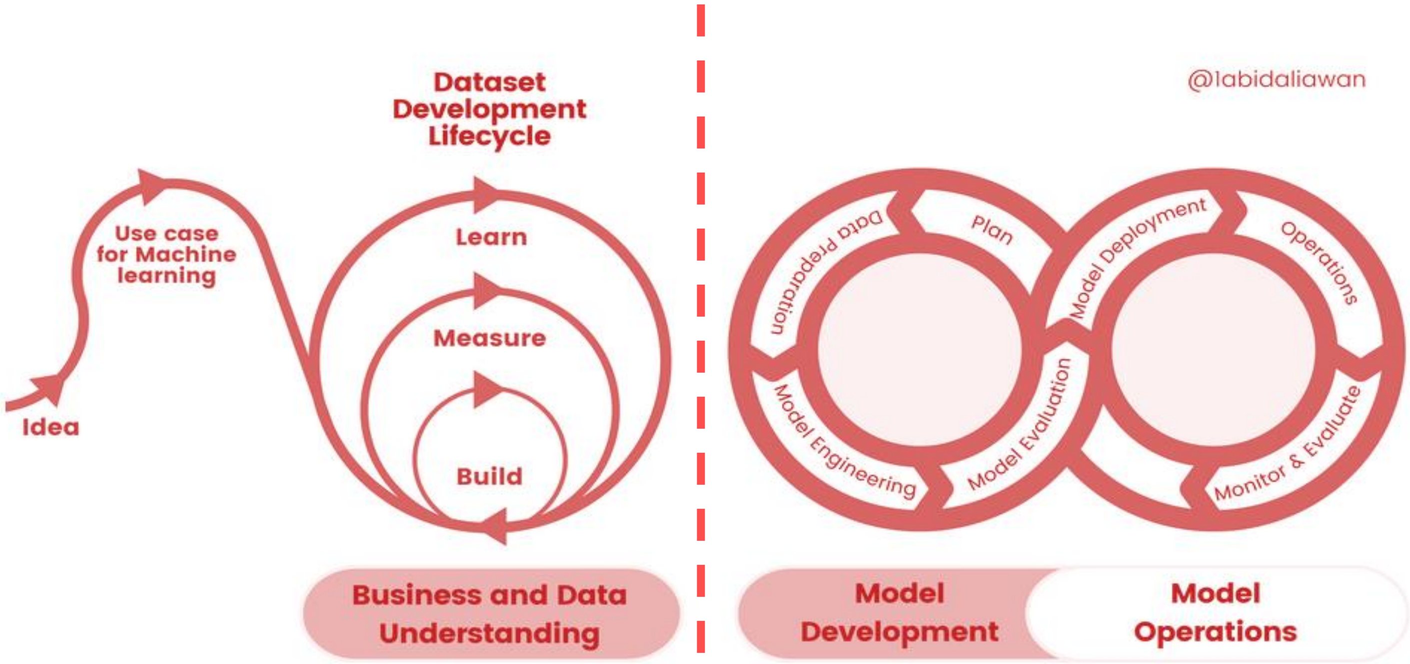
# PROBLEM STATEMENT

People have **diverse preferences** and dietary restrictions when it comes to food. Especially in India, which is an amalgamation of various cultures. How does a business account for such individual preferences and **provide personalized suggestions** based on factors like taste, ingredients, nutrition, allergies and past eating habits? And how does a **consumer discover new dishes** they might enjoy and save time in deciding what to eat? Also, with the abundance of food options available today, it can be overwhelming to make a decision, especially when **ordering food online**.

## "GOURMET GUIDE" IS THE SOLUTION

We have developed a *food recommendation model* that provides *personalized suggestions* based on *customer preferences*, driving *customer satisfaction*, loyalty, and revenue for the business and partner restaurants

# PROCESS




***Source*** *: https://www.kdnuggets.com/2022/06/making-sense-crispmlq-machine-learning-lifecycle-process.html*

# DATASET PREPARATION

What is going to be our Data?

Data of approximately 51717 Restaurants in Bangalore

What is the Source of this Data?

Zomato (Pre-scrapped data from Kaggle)

What data did we have?

- Logistics: Name, Phone Number, Address, Location, City
- About: Restaurant Type, Cuisines Offered, Approximate Cost for Two, Reviews, Ratings, Number of Votes, Dishes Liked by Users.
- URL

**Data Selection & EDA**

Name, Location, Restaurant Type, Cuisines Offered, Approximate Cost for Two, Reviews, Ratings, Dishes Liked by Users were selected.

Exploratory Data Analysis (EDA) was conducted to gain insights and understand patterns in the data.

**Data Cleaning & Text Processing**

Removed duplicate entries and handled missing values to ensure data quality. Leveraged NLP techniques to analyze and interpret restaurant reviews effectively. Employed text preprocessing methods, such as lowercasing, removing punctuation, and tokenization, to prepare the reviews for analysis.

**Data Construction**

Utilized the TF-IDF technique to quantify the importance of terms within the restaurant reviews. TF-IDF calculates a score that reflects both the frequency of a term in a review and its rarity across all reviews in the dataset. This approach enables the identification of crucial terms that carry significant meaning and contribute to the overall understanding of the reviews.

After applying NLP to the dataset, the number of instances increased from 51,717 to 80,843. To ensure computational efficiency and facilitate model training, a decision was made to build the recommendation model using a representative sample of the data. Specifically, a 15% subset of the preprocessed dataset was selected for further analysis and model development.

The chosen subset was then split into two separate sets: a training set and a test set.

# MODEL TRAINING

78% Training | 22% Testing

User in Bangalore is looking for a restaurant in "Kalyan Nagar"

Step 1:
Filtering of Restaurants based on LOCATION

**Output 1:**
List of all restaurants in "Kalyan Nagar"

Step 2:
Filtering of Restaurants based on Choice of CUISINE.

User chooses "Chinese"

**Output 2:**
List of all restaurants in "Kalyan Nagar" serving "Chinese"

Step 3:
Filtering of Restaurants based on Choice of RESTAURANT TYPE.

User chooses "Casual Dining"

**Output 3:**
List of all restaurants in "Kalyan Nagar" serving "Chinese" which is a "Casual Dining" restaurant.

Step 4:
Filtering of Restaurants based on COST.

User chooses "Rs600-Rs950"

**Output 4:**
List of all restaurants in "Kalyan Nagar" serving "Chinese" which is a "Casual Dining" restaurant with the given budget.

Step 5:
The recommendation system uses cosine similarity, which ranks restaurants by comparing the similarity of their reviews, allowing users to select a restaurant from a list of Top 5 restaurants

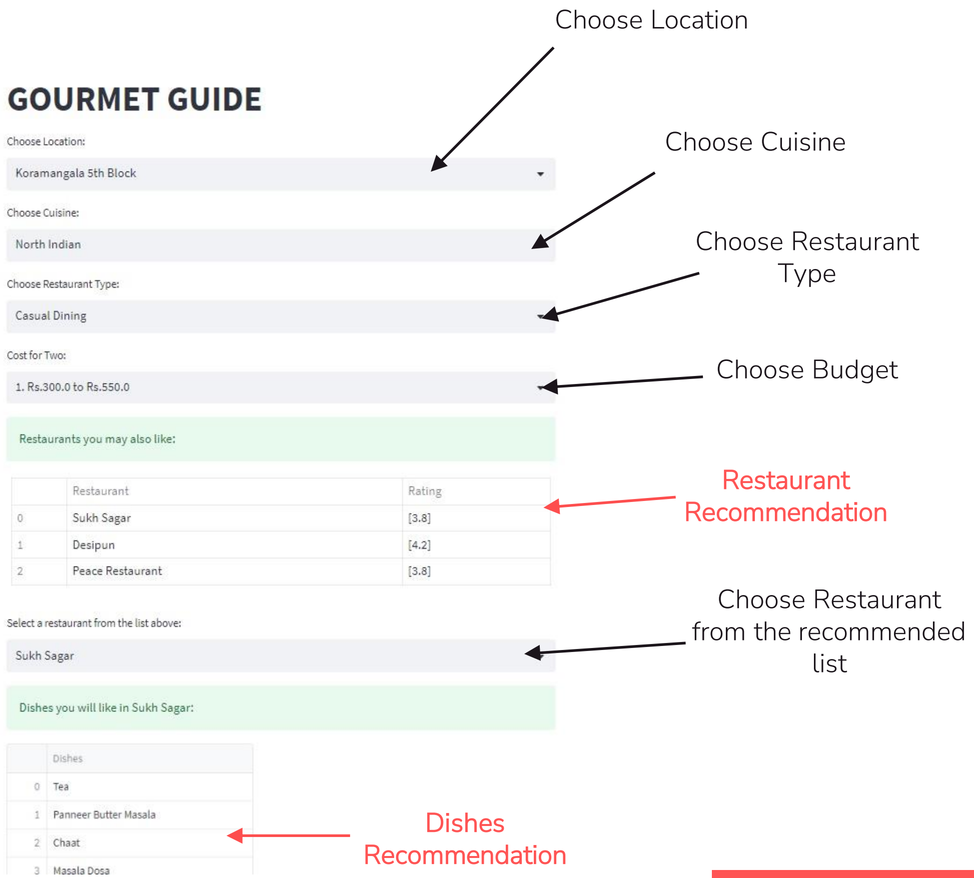
**Output 5:**
List of all Top 5 "Casual Dining" Restaurants in "Kalyan Nagar" serving "Chinese" in the given budget along with their corresponding user based rating.

Step 6:
User chooses a restaurant

**Output 6:**
List of recommended dishes for the selected restaurant based on customer reviews and preferences.

# MODEL DEPLOYMENT

"Gourmet Guide" has been successfully deployed on Streamlit, offering a user-friendly interface for exploring restaurant recommendations based on location, cuisine, and restaurant type. Users can easily navigate through different options, such as selecting the preferred location, cuisine, and restaurant type, and receive personalized recommendations tailored to their preferences.



# MEETING OUR ML SUCCESS CRITERIA

## CRITERIA MET

a. Personalization: “Gourmet Guide” accurately recommends restaurant based on users location, choice of cuisine, restaurant type, cost preference and user reviews.

b. High Accuracy: We assume a baseline accuracy of **~60%** for the recommendations in "Gourmet Guide." This assumption is based on the understanding that factors such as location, cuisine, restaurant type, and cost recommendations are **100% accurate.** However, for the cosine similarity accuracy, which measures the similarity between user preferences and restaurant reviews, we anticipate a **50-50 accuracy rate.** As we collect more data and refine our system, we aim to improve the overall accuracy to **~70% initially**, with a long-term goal of achieving **80% accuracy.** It is important to note that these accuracy figures are based on our assumptions and may vary as we gather more data and enhance the system's sophistication.

## CRITERIA TO BE MET

a. Scalability: This can be achieved by implementing distributed computing techniques, such as parallel processing or distributed frameworks like Apache Spark, to distribute the workload across multiple machines or clusters. This will allows for faster processing and increased capacity to handle a larger volume of data and user requests.

b. Real-Time Recommendations: Achieving real-time recommendations can be accomplished by implementing asynchronous processing. Instead of waiting for each individual request to complete before responding, the system can utilize distributed processing techniques to parallelize computations and effectively handle multiple requests concurrently. Additionally, it is crucial to collect data from surveys and across multiple cities in order to train our own models.

# MODEL EVALUATION

### CONFUSION MATRIX

| | Predicted & Recommended | Predicted & Not Recommended |
|---|---|---|
| Actual & Recommended | 500 (True Positive) | 500 (False Negative) |
| Actual & Not Recommended | 700 (False Positive) | 300 (True Negative) |

To evaluate the performance of our model in "Gourmet Guide," we will rely on user feedback. In our pilot study, we will ask users whether the recommended restaurants were correct or not, and their responses will be recorded as binary values (yes or no). Using this feedback, we will calculate precision, recall, and F1 score, which are common metrics for evaluating recommendation systems.

Our primary objective during the evaluation process will be to improve the accuracy of our recommendations. By analyzing user feedback and continuously refining the model based on their preferences, we aim to enhance the precision and recall values, ultimately providing more accurate and satisfying recommendations to our users.

## PRECISION

Precision measures the proportion of correctly recommended restaurants out of the total recommendations made by the system. It indicates the ability of the model to accurately identify relevant restaurants for the user.

Precision = Number of correctly recommended restaurants / (Number of correctly recommended restaurants + Number of incorrectly recommended restaurants)

Precision = 500 / (500 + 700) = 0.417

Precision = 41.7%

## RECALL

Recall measures the proportion of correctly recommended restaurants out of the total relevant restaurants in the dataset. It indicates the ability of the model to capture all the relevant restaurants for the user.

Recall = Number of correctly recommended restaurants / (Number of correctly recommended restaurants + Number of relevant restaurants not recommended)

Recall = 500 / (500 + 500) = 0.500

Recall = 50.0%

## F-1 SCORE

F-1 Scores provides a balanced evaluation of our model's performance by considering both the ability to correctly recommend relevant restaurants and the ability to avoid incorrect recommendations.

F-1 Score = 2 * (Precision * Recall) / (Precision + Recall)

F-1 Score = 2 * (0.417 * 0.5) / (0.417 + 0.5) = 0.455

F-1 Score = 45.50%

# MONITORING & MAINTENANCE

## PERFORMANCE THRESHOLDS

a. Response Time: We have a target of approximately **10 seconds** for the system to respond to user requests, ensuring a smooth and responsive user experience. If the response time exceeds this threshold, we conduct a thorough investigation to identify potential causes such as system bottlenecks or resource limitations. We optimize the system's performance by identifying areas for improvement, such as optimizing algorithms or database queries in order to reduce response time.

b. Throughput: We have a goal of handling **10 requests per minute**. To ensure that the system can handle the expected volume of user requests while maintaining the desired throughput, we conduct load testing. We regularly review and adjust the SLA for throughput as the system's usage and requirements evolve over time.

**Achieving scalability in our model will lead to notable improvements in the performance metrics. Specifically, once scalability is attained, we can expect significant enhancements in response time and throughput, thereby ensuring a smoother and more efficient user experience.**

## SCALABILITY TESTING

Scalability testing should be performed periodically to evaluate the system's capability to handle increased loads. Currently, the model operates on a sample data comprising only **15% of the entire dataset, which includes 51,000+ restaurants in Bangalore. Immediate scaling is necessary to incorporate all restaurants into our dataset.** It is important to note that the dataset is highly dynamic, with new restaurants opening, restaurants closing, and a constant generation of reviews. Therefore, the next stage of scaling will also need to be carried out swiftly, taking into account the dynamic nature of the dataset to ensure its accuracy and comprehensiveness.

**At the current stage, the model will need re-training at an increment of an addition of 10,000 restaurants or 25,000 user reviews whichever is early.**

# MEETING OUR BUSINESS SUCCESS CRITERIA

## HIGH CONVERSION RATES

To evaluate the impact of our recommendation system on conversion rates, we will track the number of users who engaged with the "Gourmet Guide" and successfully converted. By analyzing the conversion rates before and after implementing the recommendation system, we can assess the system's effectiveness in driving conversions. The **accuracy** of our model at **~60% (and growing)** plays a crucial role in this process, as it directly impacts the quality of recommendations provided to users. By ensuring the recommendations are highly accurate and tailored to individual preferences, we can enhance the user experience and increase the likelihood of conversions. Furthermore, maintaining a competitive edge in terms of our offerings compared to other platforms is essential in driving higher conversion rates.

## IMPROVED CUSTOMER RETENTION

By analyzing customer behavior data, including metrics like repeat visits and usage frequency, we can assess the impact of our recommendation system on customer retention. The recommendation system plays a vital role in minimizing customer decision time by delivering personalized and relevant suggestions within a **10 second response time**, which will further decrease as we achieve scale. If users are returning to the platform more frequently or engaging with it on a regular basis, it suggests that the recommendation system has successfully improved customer retention. The ability to quickly deliver tailored recommendations enhances the overall user experience, leading to increased customer satisfaction and loyalty.

# MEETING OUR ECONOMIC SUCCESS CRITERIA

## REVENUE GROWTH

Total Addressable Market (TAM): To estimate the TAM, we consider the total number of urban and rural households in India and believe that approximately 15% of them eat out.

Urban households + Rural Households = 130mn + 172mn = 302mn

Total TAM = 302mn x 15% = ~45.3 million households

a. <u>User Subscription For Premium Features</u>: Our goal to implement subscription plans at Rs.500/ year for premium features, which will result in remarkable revenue growth. Users can enjoy exclusive access to curated recommendations, priority booking options, and attractive deals at our partner restaurants.

Revenue = Total TAM x Adoption Rate x Rs. 500
= 45mn x 5% (reasonable estimate) x 500
= ~Rs.112mn

b. <u>Partnerships with Restaurants</u>: Our strategic partnerships with restaurants will significantly contribute to our revenue growth. By offering restaurants the opportunity to be prominently featured in recommendations or highlighted as preferred options for a fee, we aim to create mutually beneficial collaborations. Restaurants benefit from increased visibility and customer engagement, while we generate additional revenue through showcasing their establishments at the top of search results, providing comprehensive restaurant profiles, and offering targeted advertising opportunities. An average fee charged to restaurants for premium placement and advertising opportunities will be Rs10,000/ year.

Revenue = Total number of Restaurants in Bangalore x Adoption Rate x Rs. 10,000
= 51000 x 50% (reasonable estimate) x 10,000
= ~Rs.255mn

Total Estimated Revenue = 112 + 255 = Rs.367mn

# MEETING OUR ECONOMIC SUCCESS CRITERIA

## COST REDUCTION AND EFFICIENCY

Our other key objective is to enhance operational efficiency and cost-effectiveness by leveraging our recommendation system to optimize business operations. To achieve this, we shall streamline processes such as restaurant selection and resource allocation, resulting in reduced costs and improved efficiency. By forming partnerships with restaurant chains and onboarding multiple restaurants at reduced expenses, we can benefit from cost savings and leverage mutual marketing opportunities. Additionally, our location-based approach allows us to onboard neighboring restaurants simultaneously, maximizing efficiency and minimizing operational expenses. Through these strategies, we aim to drive cost reduction, improve efficiency, and maintain a competitive edge in the market.

# DID WE MEET OUR OBJECTIVE?

## MAXIMIZE REVENUE

By implementing our "Gourmet Guide" platform, we are confident in our ability to achieve our revenue maximization goals. Our strategic approach focuses on fostering repeat business by incentivizing customer loyalty and utilizing data-driven insights. Through thorough analysis of customer behavior and preferences, we have successfully generated highly personalized recommendations. This personalized approach has resulted in a significant boost in sales and revenue.

Based on our estimations, we anticipate generating approximately Rs. 367 million in revenue within the first year of operations.

## MINIMIZE DECISION TIME FOR CUSTOMERS

With our platform, "Gourmet Guide," we are dedicated to reducing customer decision time through the utilization of advanced algorithms and customer data analysis. By offering efficient and tailored recommendations, we enable customers to swiftly make their food choices, increasing the probability of placing orders without prolonged deliberation.

Currently, our response time stands at an impressive 10 seconds, and we are continuously striving to further decrease it.

# YES!!!

# CONSTRAINTS & ASSUMPTIONS

## CONSTRAINTS WITH "GOURMET GUIDE"

a. Stagnant Data Availability: The model relies on historical data and did not collect real-time data. This means that the available data may not reflect recent changes in the restaurant industry, such as closures and new openings.

b. Dynamic Nature of Restaurants: Restaurants are dynamic in nature, and their operability can change over time. Some restaurants may start off strong and then fade away, while others may initially be small and eventually grow. This dynamic nature of restaurants is not captured in the model.

c. Variability in Delivery and Dine-in Options: Delivery and dine-in options offered by restaurants can vary over time. However, the model assumes a static representation of these options and does not consider changes in the availability of delivery or dine-in services.

## ASSUMPTIONS IN MAKING THE MODEL

a. No Changes in Restaurants: The model assumes that there will be no changes in the restaurants included in the dataset. It does not account for new restaurants opening or existing ones closing down.

b. Continued Operation of Existing Restaurants: The model assumes that the existing restaurants in the dataset will continue to operate as per the available data. It does not consider the possibility of changes in ownership, management, or quality.

c. Reliance on Historical Ratings and Reviews: The recommendations provided by the model are based on the ratings and reviews of customers as per the earlier data. The model assumes that the restaurants will maintain their quality based on these ratings and reviews, without considering potential changes in service or food quality over time.

# REFERENCES/ LITERATURE


- Content-Based Recommendation Systems by Michael J. Pazzani, Daniel Billsus
- Restaurant recommender system based on sentiment analysis by Elham Asani, Hamed Vahdat-Nejad, Javad Sadri
- Yelp Food Recommendation System by Sumedh Sawant, Gina Pai
- Dataset from: https://www.kaggle.com/code/chirag9073/zomato-recommendation-system/input


# THANK YOU

## PRESENTED BY: GROUP - 6

12220064: Charanjeet Singh

12220028: Pooja Nilesh Doshi

12220067: Snigdha Debashis Bhattacharjee

12220047: Vinayak Dave