MealMaster: A Recipe Bot Recommendation System

Group7

Ahmed Abdelaal

Motaz Habib

Shahd Mohamed

Yousef Mohamed

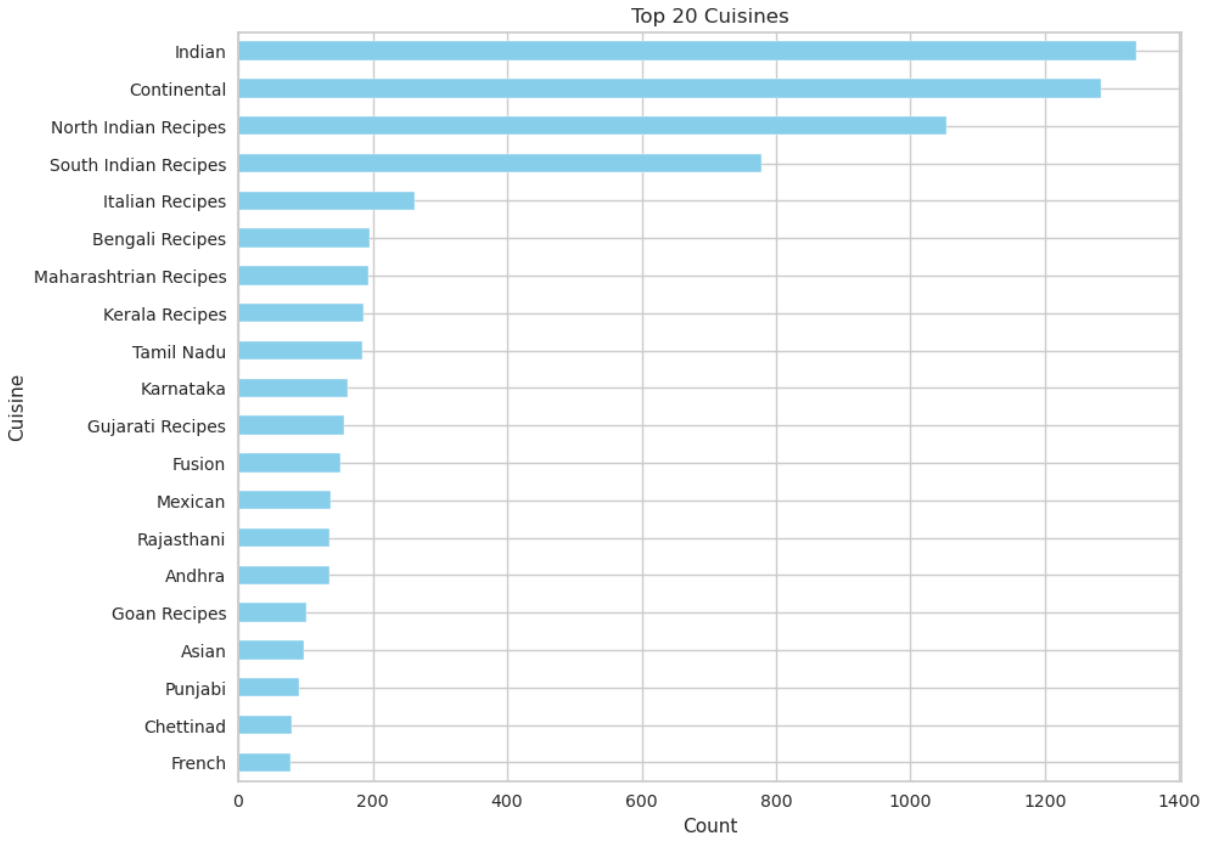
1. Problem Formulation:

Our project aims to develop a chatbot that provides personalized recipe recommendations. This chatbot will use two crucial factors to generate suggestions: the user's preferred cuisine and available ingredients. To accomplish this, we will utilize clustering to categorize similar recipes and classification to identify the most suitable category based on the user's preferences. All this will be part of a chatbot that's easy and convenient for users to find new recipes.

2. Data and Data Preprocessing:

Our project uses a dataset from Kaggle, which has over 8,000 recipes, most of them being Indian dishes. This dataset gives us many details about each recipe, like how to cook it, the ingredients needed, how long it takes, who wrote the recipe, and how it's rated. We can use this information to compare different recipes, and group similar ones together.

We began our project by performing an initial analysis of the dataset, examining the distribution and frequency of each feature as depicted in (*Figure 1*).
We saw that some types of cuisines didn't have many recipes, and there was a lot more data for 'Indian' cuisine than for other types of cuisine. This means our data was not balanced.



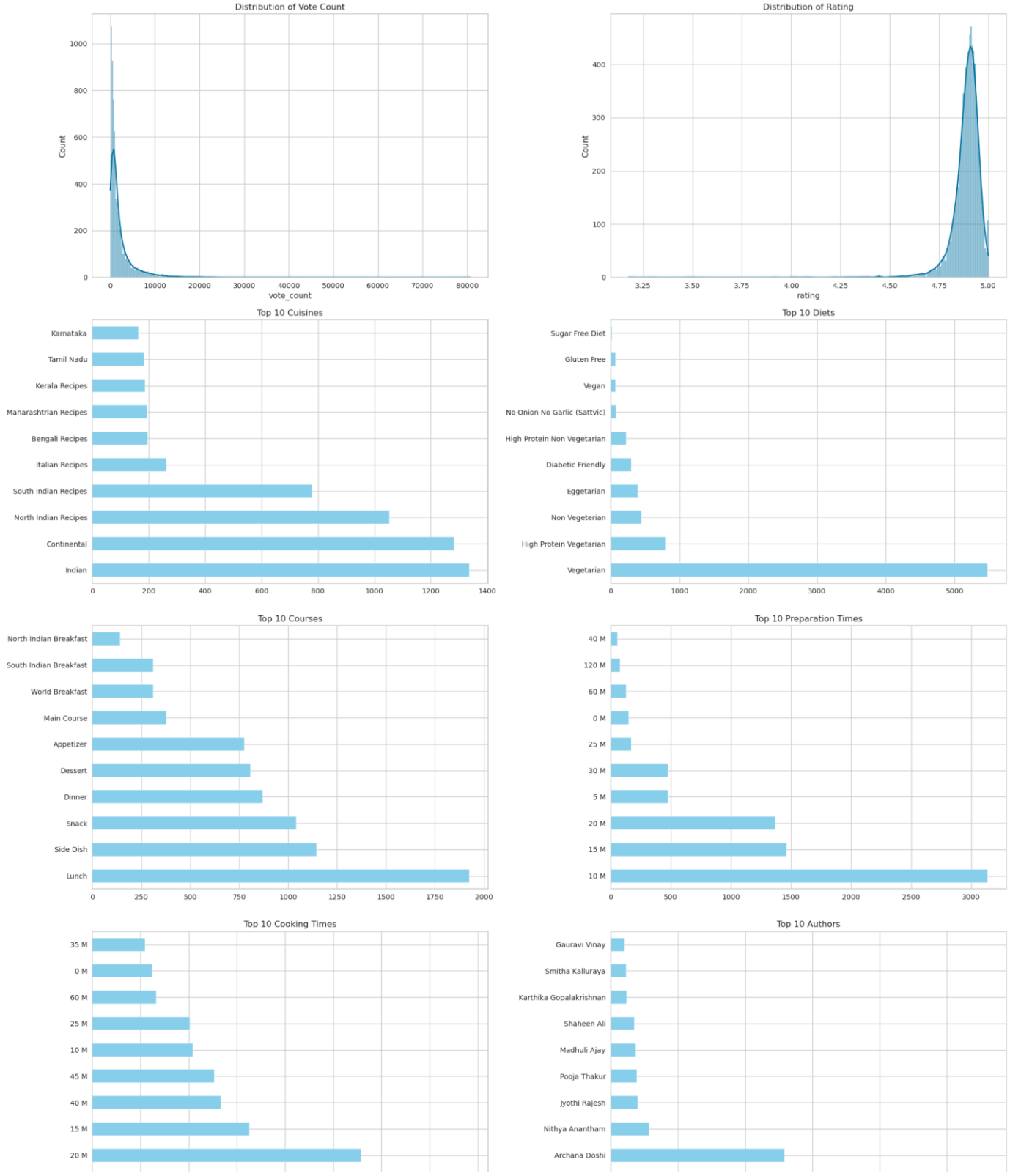
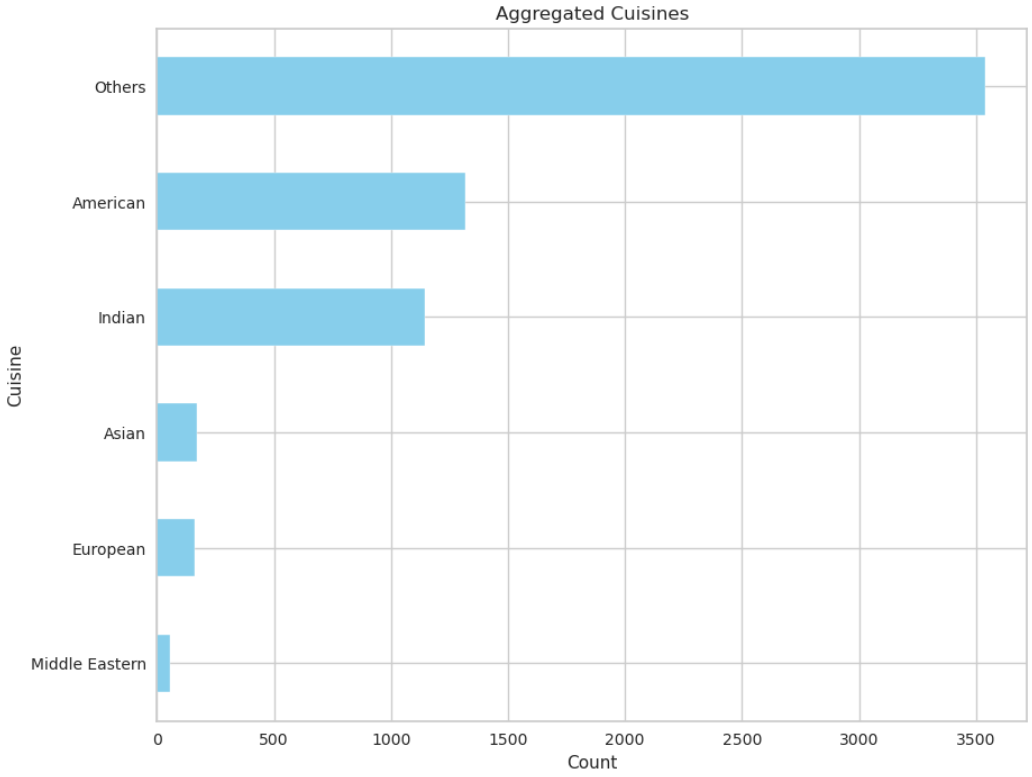

Distribution of Vote Count
Count
vote_count
Distribution of Rating
Count
rating
Top 10 Cuisines
Karnataka
Tamil Nadu
Kerala Recipes
Maharashtrian Recipes
Bengali Recipes
Italian Recipes
South Indian Recipes
North Indian Recipes
Continental
Indian
Top 10 Diets
Sugar Free Diet
Gluten Free
Vegan
No Onion No Garlic (Sattvic)
High Protein Non Vegetarian
Diabetic Friendly
Eggetarian
Non Vegeterian
High Protein Vegetarian
Vegetarian
Top 10 Courses
North Indian Breakfast
South Indian Breakfast
World Breakfast
Main Course
Appetizer
Dessert
Dinner
Snack
Side Dish
Lunch
Top 10 Preparation Times
40 M
120 M
60 M
0 M
25 M
30 M
5 M
20 M
15 M
10 M
Top 10 Cooking Times
35 M
0 M
60 M
25 M
10 M
45 M
40 M
15 M
20 M
Top 10 Authors
Gauravi Vinay
Smitha Kalluraya
Karthika Gopalakrishnan
Shaheen Ali
Madhuli Ajay
Pooja Thakur
Jyothi Rajesh
Nithya Anantham
Archana Doshi

Figure 1

Merging Cuisines:

To make things easier, we put cuisines into bigger groups. We also had some missing information in our data. To handle this, we removed the parts of the data that had missing info and we also



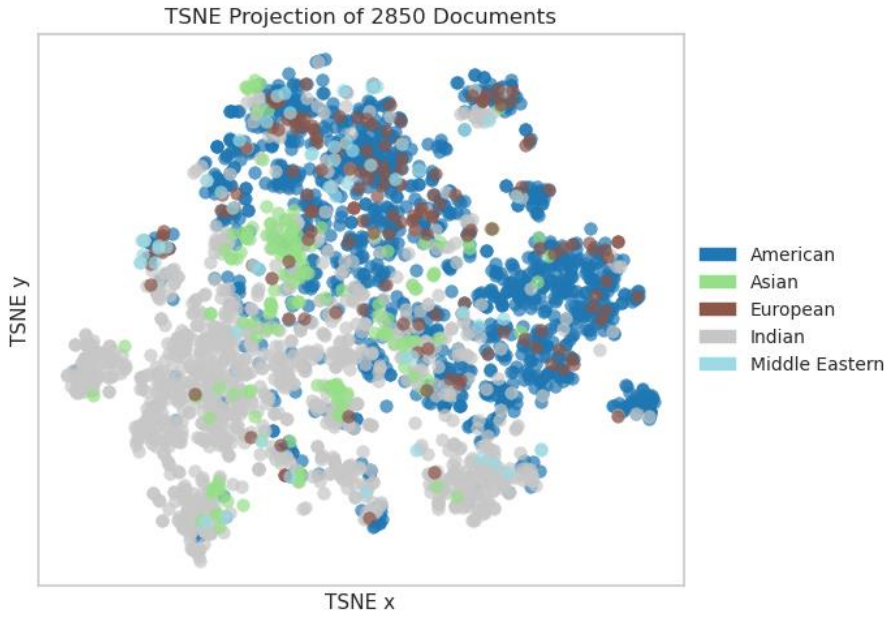
removed the “Others” from the Data.

3. Feature Engineering:

To transform our data for effective modeling, we utilized TF-IDF (Term Frequency-Inverse Document Frequency) and Word2Vec. These strategies convert textual information into numerical vectors.

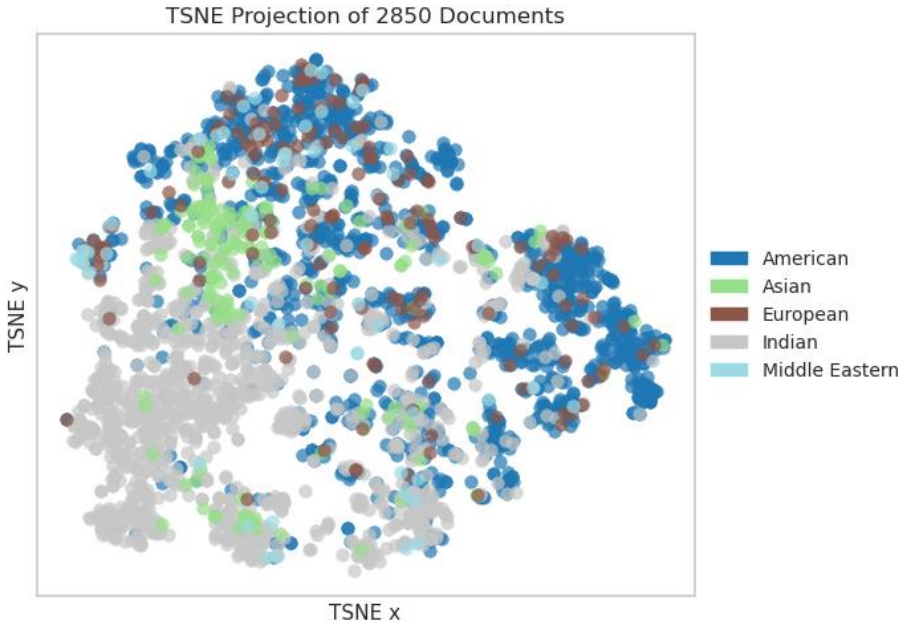
1. TF-IDF:

In our project, we employed TF-IDF this technique evaluates the importance of a word in a document relative to a corpus, allowing us to discern the key ingredients and processes in each recipe and use this information in our recommendation system.



2. Word2Vec:

In addition to TF-IDF, we also used Word2Vec, a method that creates word embeddings by analyzing the context in which words appear. This technique allows us to capture semantic relationships between words.

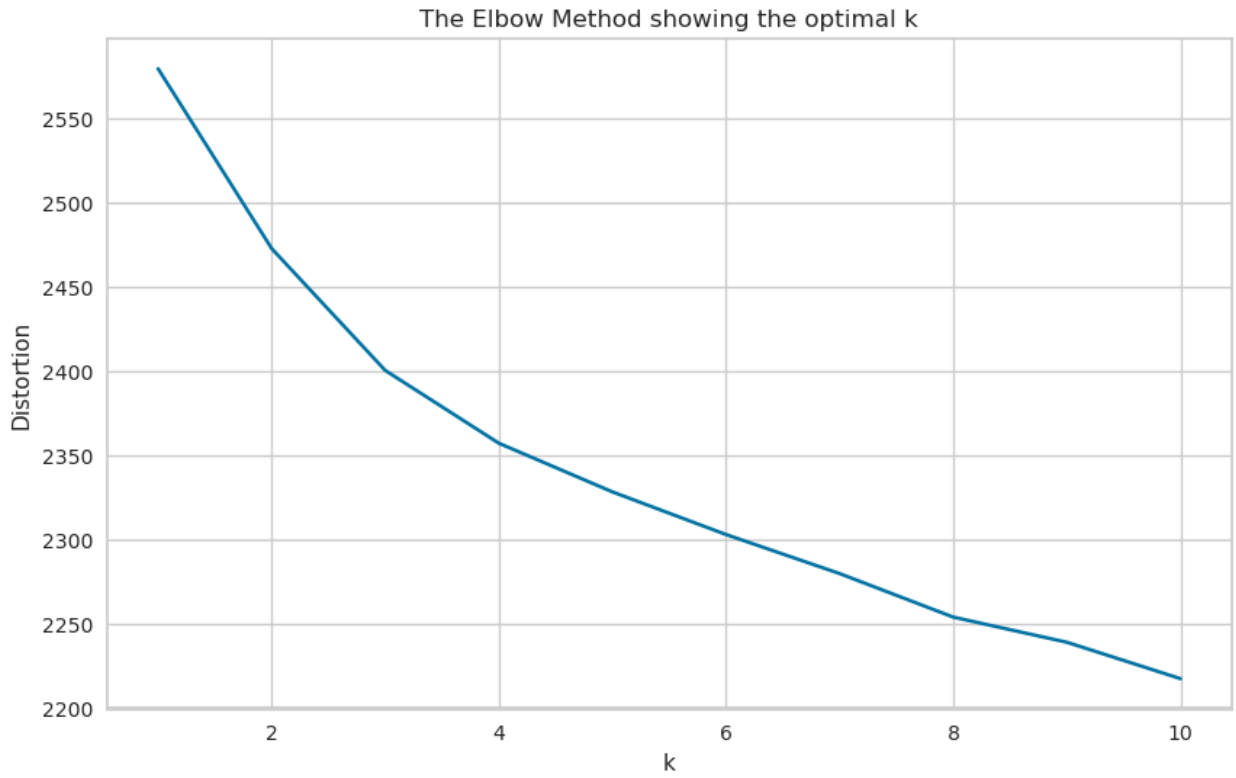


We noticed that TF-IDF provided a clearer separation of our data compared to Word2Vec. While Word2Vec captures semantic relationships and is excellent for understanding context, it was the TF-IDF method that excelled in distinguishing and categorizing our cuisines more distinctly.

4. Modeling and Model Evaluation:

1. Clustering:

We used the k-means clustering algorithm to group our recipes, specifically focusing on their ingredient profiles. This approach allowed us to identify and categorize similar recipes, offering a structured way to make recipe recommendations based on ingredient similarities.



By using the elbow method, we found that the best way to group them was into four main clusters.

3. Recommendation System:

Procedure:

1- Recipe Profiling:

- Features Utilized: Cuisine type, course, diet, and ingredients.
- Data Transformation: One-hot encoding is applied to categorical metadata like cuisine type, course, and diet. The ingredients are vectorized using Term Frequency-Inverse Document Frequency (TF-IDF) to capture their importance.
- Profile Creation: The one-hot encoded and TF-IDF vectorized features are horizontally stacked to create a comprehensive profile for each recipe.

2- Data Preparation:

- The DataFrame's index is reset to ensure a sequential order, which is essential for accurate indexing and retrieval of recipes.

3- Similarity Computation:

- Method: Cosine similarity, a metric that measures the cosine of the angle between two non-zero vectors, is computed between the recipe profiles. A value closer to 1 indicates a higher similarity.
- Outcome: A matrix of cosine similarities between recipes is generated.

4- Recommendation Algorithm:

- A function, recommend_recipes_content_based, is designed to recommend similar recipes:
    - The provided recipe title is standardized (converted to lowercase and stripped of white spaces).
    - The index of the specified recipe in the DataFrame is identified.
    - Cosine similarities for the given recipe against all others are retrieved.
    - Recipes are sorted based on their similarity scores.
    - The cluster or category (e.g., dessert, main course) of the given recipe is identified. This acts as an additional filtering criterion.
    - The function recommends recipes that belong to the same cluster as the input and excludes the input recipe itself from the recommendations.
    - The recommended recipes are then returned in the form of a DataFrame.

5- Test Case:

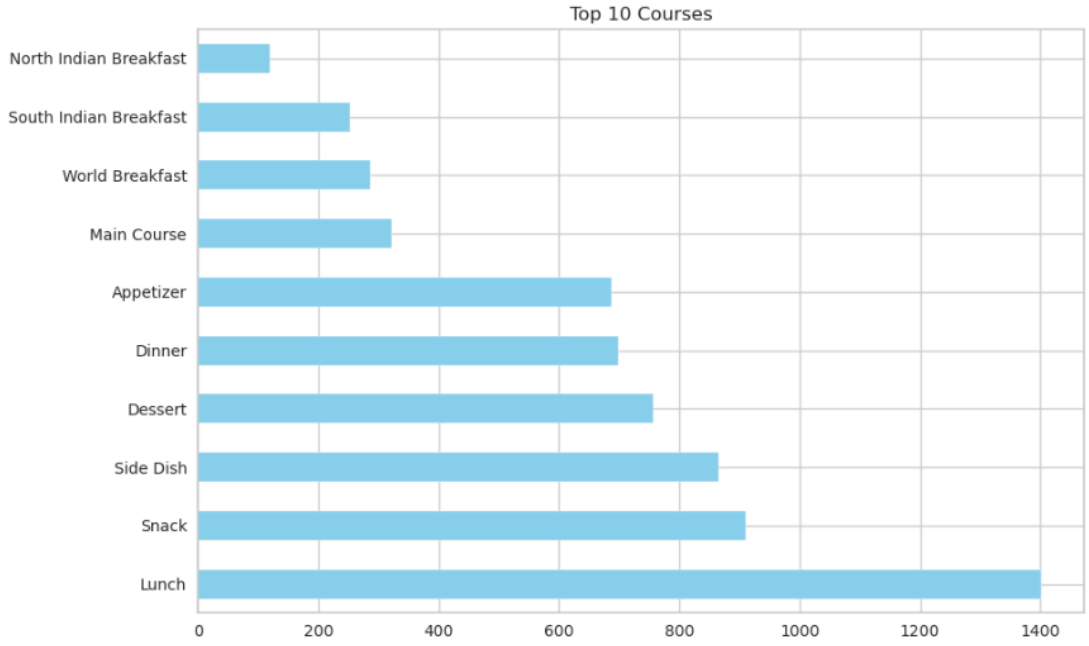
- The recommendation system is tested with the recipe titled "Pineapple Upside Down Cake Recipe". The system successfully recommends five similar recipes from the dataset.

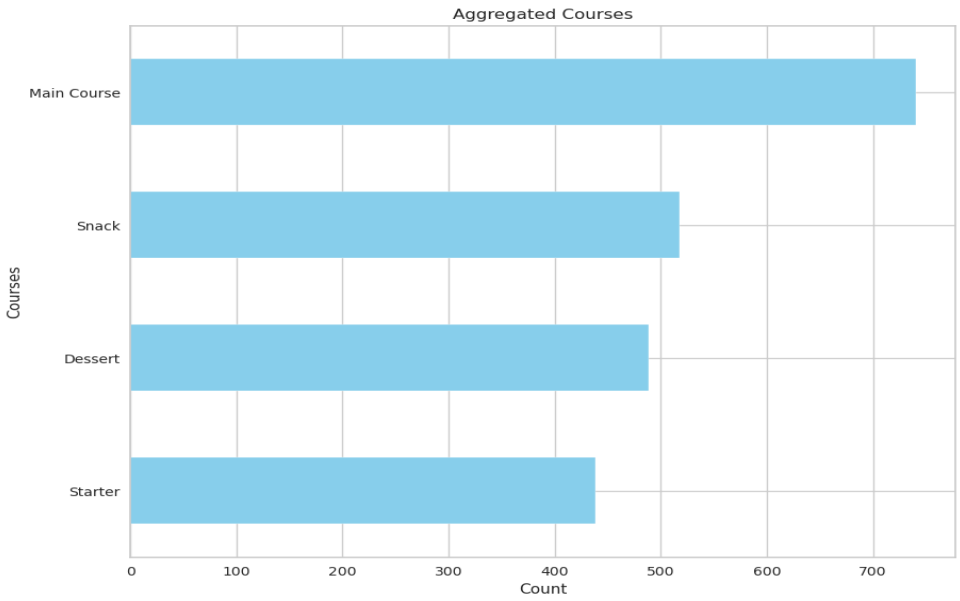
	recipe_title	description	cuisine	course	diet	prep_time	cook_time	ingredients	instructions	category	cuisine_agg	course_agg	ingredients_cleaned	cluster_ingredients
21	Eggless Chocolate Mug Cake Recipe - Instant Mi...	Chocolate Mug Cake Recipe is a quick-fix, inst...	continental	Dessert	Vegetarian	10	1	butter unsalted|cocoa powder|sugar|all purpose...	To begin making the Eggless Chocolate Mug Cake...	Cake Recipes	American	Dessert	butter unsaltedcocoa powdersugarall purpose fl...	0
940	Chocolate Cherry Cake Recipe	Chocolate Cherry Cake Recipe is from Scratch a...	continental	Dessert	Vegetarian	45	15	whole eggs|all purpose flour maida|cocoa powde...	To begin with the Chocolate Cherry Cake From ...	Cake Recipes	American	Dessert	whole eggsall purpose flour maidacocoa powderm...	0
1324	Carrot Cake Parfait With Custard And Strawberr...	Moist chunks of carrot cake (with whole wheat)...	continental	Dessert	Vegetarian	10	40	sugar|brown sugar demerara sugar|sunflower oil...	To begin making the Carrot Cake Parfait With C...	Dessert Recipes	American	Dessert	sugarbrown sugar demerara sugarsunflower oilwh...	0
1986	Eggless Lemon Pound Cake Recipe	The Old Fashioned is a buttery cake perfec...	continental	Dessert	Vegetarian	5	45	butter unsalted|caster sugar|all purpose flour...	To begin preparing the Lemon Pound Cake, prehe...	Cake Recipes	American	Dessert	butter unsaltedcaster sugarall purpose flour m...	0
2054	Delicious Strawberry Tea Cake Recipe	The Wholesome Strawberry Cake is just the kind...	continental	Dessert	Vegetarian	20	50	butter salted|brown sugar demerara sugar|whole...	To begin making the 50% Whole-Wheat Strawberry...	Cake Recipes	American	Dessert	butter saltedbrown sugar demerara sugarwhole w...	0

4. Classification:

In our project, we tried three methods to make our Recipe Bot work best: Support Vector Machines (SVM), Logistic Regression, and Random Forest (RF). We tested each one to see which gave the best recipe suggestions.

Before we started predicting with methods like SVM, Logistic Regression, and Random Forest, we grouped similar recipes together. This made our Recipe Bot give better and more accurate suggestions.





In evaluating the performance of our models, we obtained the following accuracy scores:

1. Logistic Regression: 60.96%
2. Support Vector Machines (SVM): 59.13%
3. Random Forest: 58%

Among the three models, Logistic Regression yielded the highest accuracy, closely followed by SVM, with Random Forest slightly lagging behind.

5. Error Analysis:

1. **Logistic Regression**:

Accuracy: 61%

Pros: Achieved the highest accuracy among the three. Consistently good performance across all classes, especially for Class 0.

Cons: Some challenges with Classes 2 and 3, but still relatively strong.

2. **SVM** (Support Vector Machines):

Accuracy: 59.13%

Pros: Good performance, also for Class 0. Close in accuracy to Logistic Regression.

Cons: Faced challenges with Classes 2 and 3, particularly in recall.

3. **Random Forest**:

Accuracy: 58%

Cons: Achieved the lowest accuracy among the three models. Had difficulties with Class 3 in both precision and recall.

Logistic Regression is the best model for this dataset. It achieved the highest accuracy and showed consistently good performance across all classes. While SVM and Random Forest were competitive, they lagged slightly behind in terms of overall accuracy and faced more pronounced challenges with certain classes.

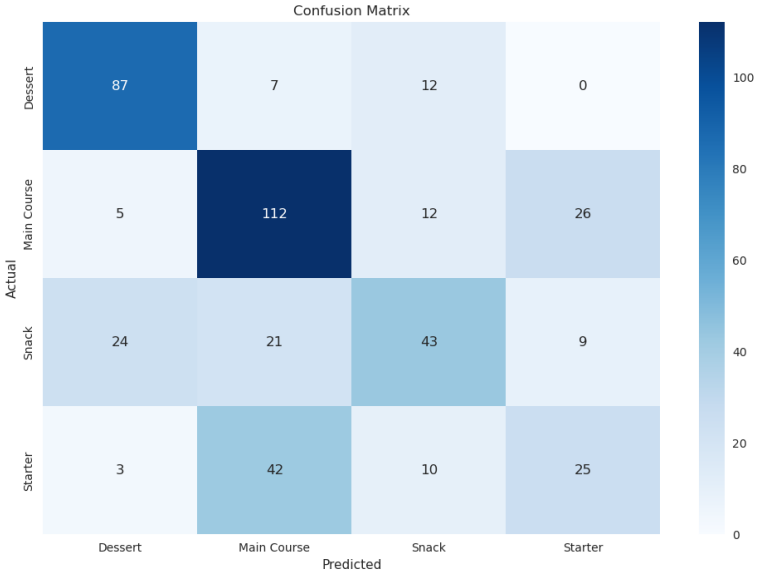
	precision	recall	f1-score	support
Dessert	0.73	0.82	0.77	106
Main Course	0.62	0.72	0.66	155
Snack	0.56	0.44	0.49	97
Starter	0.42	0.31	0.36	80
accuracy			0.61	438
macro avg	0.58	0.57	0.57	438
weighted avg	0.59	0.61	0.60	438

Logistic Regression

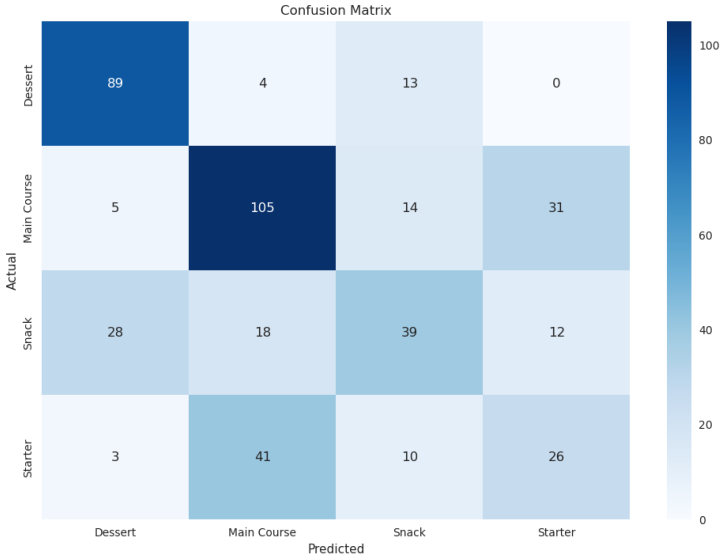
	precision	recall	f1-score	support
Dessert	0.71	0.84	0.77	106
Main Course	0.62	0.68	0.65	155
Snack	0.51	0.40	0.45	97
Starter	0.38	0.33	0.35	80
accuracy			0.59	438
macro avg	0.56	0.56	0.56	438

	precision	recall	f1-score	support
Dessert	0.71	0.76	0.74	106
Main Course	0.61	0.72	0.66	155
Snack	0.50	0.35	0.41	97
Starter	0.35	0.31	0.33	80
accuracy			0.58	438
macro avg	0.54	0.54	0.53	438
weighted avg	0.56	0.58	0.56	438

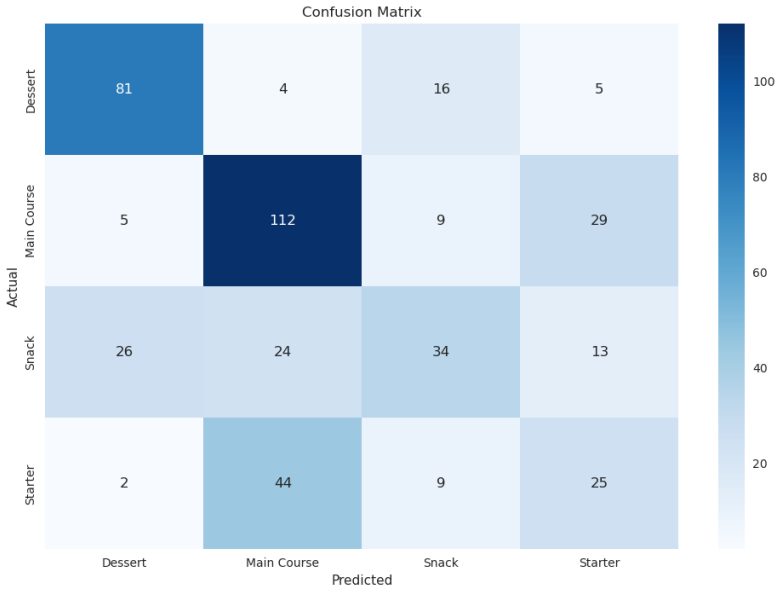
Random Forrest



Logistic Regression



SVM



Random Forest

Examples of the performance differences between models:

	ingredients	Actual Course	Predicted Course
2	butter unsalted|caster sugar|whole eggs|all pu...	Dessert	Snack
6	mango ripe|tomato|onion|rajma large kidney bea...	Snack	Main Course
7	black eyed beans lobia|rice|carrots gajjar|gre...	Starter	Main Course
15	potatoes aloo|garlic|fresh thyme leaves|salt	Snack	Dessert
16	chicken|onion|tomato|green chillies|bay leaves...	Snack	Main Course
...	...	...	...
428	all purpose flour maida|red chilli powder|ging...	Snack	Main Course
429	sabudana tapioca pearls|milk|sugar|cardamom po...	Main Course	Snack
431	mango ripe|fresh cream|milk|condensed milk	Snack	Main Course
435	spinach leaves palak|coriander dhania leaves|p...	Main Course	Starter
436	brown sugar demerara sugar|cocoa powder|milk|m...	Snack	Dessert

Logistic Regression

	ingredients	Actual Course	Predicted Course
2	butter unsalted|caster sugar|whole eggs|all pu...	Dessert	Snack
6	mango ripe|tomato|onion|rajma large kidney bea...	Snack	Main Course
7	black eyed beans lobia|rice|carrots gajjar|gre...	Starter	Main Course
15	potatoes aloo|garlic|fresh thyme leaves|salt	Snack	Dessert
16	chicken|onion|tomato|green chillies|bay leaves...	Snack	Main Course
...	...	...	...
430	tortillas|iceberg lettuce|onion|tomato|green b...	Snack	Starter
431	mango ripe|fresh cream|milk|condensed milk	Snack	Main Course
434	bermuda grass powder|ginger|salt|black pepper ...	Starter	Main Course
435	spinach leaves palak|coriander dhania leaves|p...	Main Course	Starter
436	brown sugar demerara sugar|cocoa powder|milk|m...	Snack	Dessert

SVM

	ingredients	Actual Course	Predicted Course
2	butter unsalted|caster sugar|whole eggs|all pu...	Dessert	Snack
4	shrimps|butter salted|orange color|lemon juice...	Main Course	Dessert
10	sabudana tapioca pearls|shallots|mixed vegetab...	Snack	Starter
15	potatoes aloo|garlic|fresh thyme leaves|salt	Snack	Dessert
16	chicken|onion|tomato|green chillies|bay leaves...	Snack	Main Course
...	...	...	...
429	sabudana tapioca pearls|milk|sugar|cardamom po...	Main Course	Snack
430	tortillas|iceberg lettuce|onion|tomato|green b...	Snack	Main Course
431	mango ripe|fresh cream|milk|condensed milk	Snack	Main Course
435	spinach leaves palak|coriander dhania leaves|p...	Main Course	Starter

Random Forrest

Conclusion:

1. **Challenges with Class 3**: All three models struggled to accurately predict Class 3, as evidenced by both low precision and recall values. This suggests that Class 3 instances might be hard to distinguish from other classes, possibly due to overlapping features or insufficient training data.

2. **Class Imbalance Impact**: The varying 'support' values (number of instances) for each class is because of an imbalanced dataset. When there's a class imbalance, models tend to favor the majority class, which can lead to poorer performance on minority classes. This could be a reason why certain classes, especially Class 3, are not being predicted as effectively.

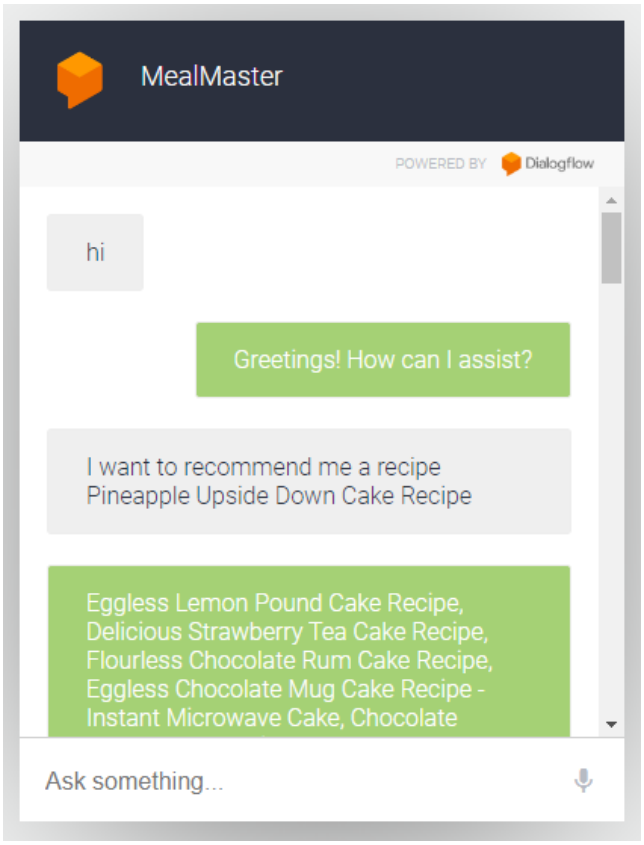
3. **Bad Recall for Certain Classes**: Besides Class 3, some models also had difficulty with Class 2. This means that a significant number of true instances for these classes were being missed or wrongly classified.

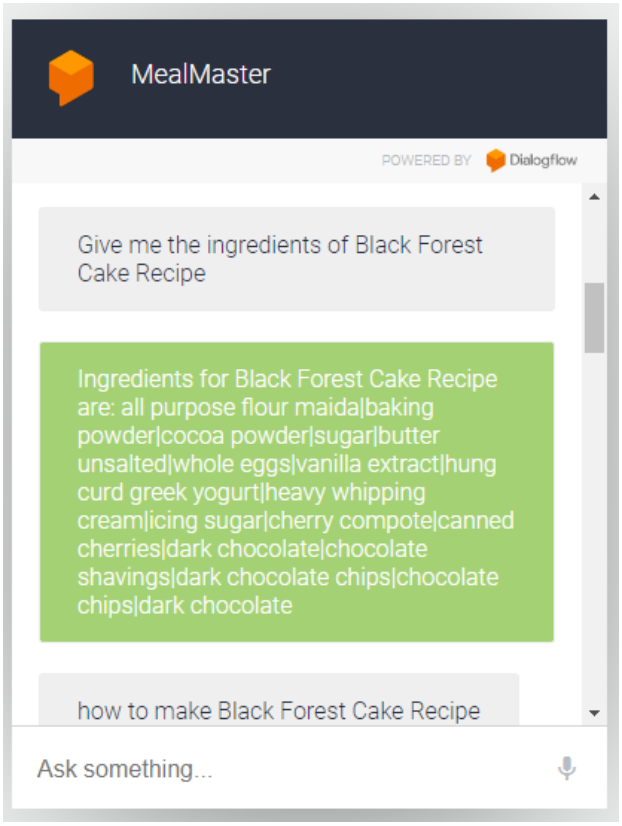
4. **Precision Challenges**: While some classes had decent recall, their precision was lacking. This means that while the models were capturing a good number of true positive instances, they were also misclassifying other instances into these classes, leading to false positives.

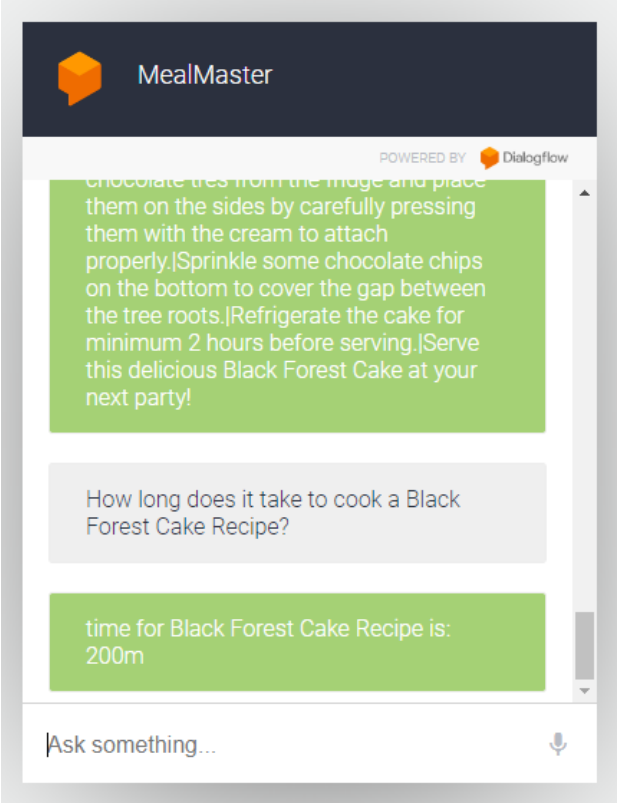
5. **Overall Accuracy**: The models didn't reach 65% accuracy, so there's more work to do. Even though we improved the data features, tweaking model settings and looking at dataset issues might help boost results.

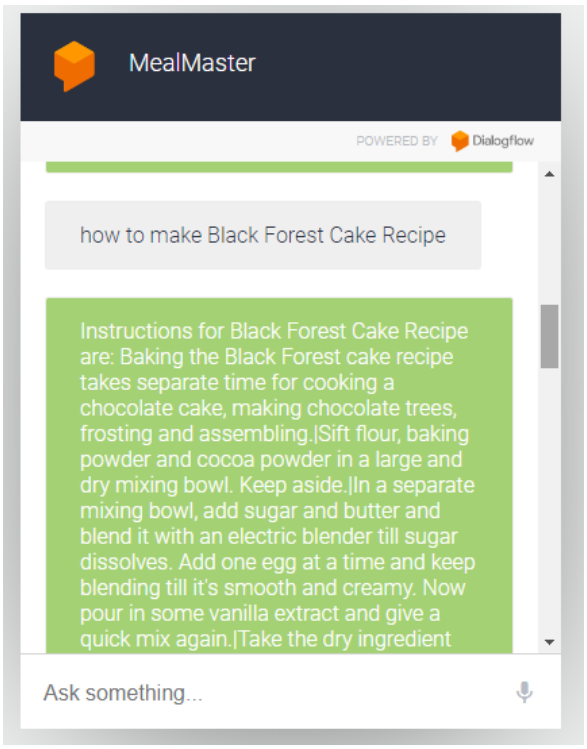
6. Chatbot:

These are screenshots from our chatbot.


MealMaster
POWERED BY Dialogflow
hi
Greetings! How can I assist?
I want to recommend me a recipe Pineapple Upside Down Cake Recipe
Eggless Lemon Pound Cake Recipe, Delicious Strawberry Tea Cake Recipe, Flourless Chocolate Rum Cake Recipe, Eggless Chocolate Mug Cake Recipe - Instant Microwave Cake, Chocolate
Ask something...


MealMaster
POWERED BY Dialogflow
Give me the ingredients of Black Forest Cake Recipe
Ingredients for Black Forest Cake Recipe are: all purpose flour maida|baking powder|cocoa powder|sugar|butter unsalted|whole eggs|vanilla extract|hung curd greek yogurt|heavy whipping cream|icing sugar|cherry compote|canned cherries|dark chocolate|chocolate shavings|dark chocolate chips|chocolate chips|dark chocolate
how to make Black Forest Cake Recipe
Ask something...





The weakness of the chatbot is that we should enter the exact recipe name, and any misspelling lead to the recipe not found.