University at Buffalo

Department of Computer Science and and Engineering

CSE 587 - Data Intensive Computing

Project Phase 3

Restaurant Recommendation System

NAME	UBIT	PERSON NUMBER
Shreya Dhareshwar	shreyadh	50471594
Rishab Aggarwal	rishabag	50475990

1] Problem Statement:

"We will use the restaurant dataset to create a Restaurant Recommendation System and use a data driven solution to solve the problem of finding suitable locations to open restaurants"

2] Code Documentation and Execution Instructions:

The code comments are contained in the code. The 'README.md' file, which is located at the top directory level of our submission, contains execution instructions, and we have also included the'requirements.txt' file.

3] Phase 2 Model usage:

We used 6 models in Phase 2 for our recommendation system. A summary is as follows:

1. Linear Regression
   The goal of linear regression is to find the best linear equation for the data and use it to make predictions.
2. Logistic Regression
   Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set.
3. KNN
   K-Nearest Neighbors (KNN) is a popular algorithm used in recommendation systems for item-based collaborative filtering.
4. Matrix Factorization
   Matrix Factorization is a technique for reducing the dimensionality of a large matrix by decomposing it into two or more matrices of lower dimension.
5. Random Forest
   It is an ensemble learning method that creates multiple decision trees and combines their predictions to obtain a more accurate and stable prediction.
6. Bag of Words
   A bag-of-words model is a technique used to represent text as a collection of the frequency of individual words that appear in a document

The metrics of the models are as follows:

Model	Metrics
Linear Regression	MSE: 0.1199, R-Squared: 0.2342
Logistic Regression	Accuracy Score: 0.85898
KNN	Accuracy score: 0.912
Matrix Factorization	MSE: 12.16268
Random Forest	Accuracy score: 0.9279

The goal of phase 3 was to demonstrate a user application with real-world applications based on the outcomes of phase 2. As a result, we developed a user application that takes user input regarding restaurant features and forecasts the best site for the restaurant's debut based on that data.

We utilized the KNN algorithm to accomplish this. Because of its proximity-based approach, non-parametric nature, interpretability, flexibility in feature engineering, and adaptability to changing data, the KNN algorithm is the ideal model for identifying the optimal site for launching a restaurant. KNN can successfully discover similar places for appropriate restaurant placement by taking into account the features of highly rated restaurants. Its ability to handle diverse datasets and provide interpretable results makes it a reliable choice for this application, allowing entrepreneurs and investors to make data-driven decisions with confidence.

The chosen value for k in this project is 5. By setting the k-value to 5, the system aims to strike a balance between considering enough neighboring data points to make accurate predictions and avoiding overfitting or underfitting the model. This value is chosen based on experimentation and evaluation of the model's performance on the given dataset.

We have the following columns in our dataset:

```
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 17 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   url           51717 non-null  object
 1   address       51717 non-null  object
 2   name          51717 non-null  object
 3   online_order  51717 non-null  object
 4   book_table    51717 non-null  object
 5   rate          43942 non-null  object
 6   votes         51717 non-null  int64
 7   phone         50509 non-null  object
 8   location      51696 non-null  object
 9   rest_type     51490 non-null  object
 10  dish_liked    23639 non-null  object
 11  cuisines      51672 non-null  object
 12  cost for two  51371 non-null  object
 13  reviews_list  51717 non-null  object
 14  menu_item     51717 non-null  object
 15  food_type     51717 non-null  object
 16  city          51717 non-null  object
dtypes: int64(1), object(16)
```

By developing a user interface we have made it highly convenient for the user to enter values for these columns or upload the dataset manually.

4] **Analysis:**

Based on the dataset and our code, The accuracy score of the trained model comes out to be 0.9101. It also identifies the best location for a restaurant based on the highest predicted success probability. It generates a scatter plot to visualize the success probability of restaurants in different locations. The x-axis represents the 'votes' feature, the y-axis represents the 'cost for two' feature, and the color represents the predicted success probability.
The best location (with the highest predicted probability) is marked with a red 'X' on the scatter plot.The colorbar provides a reference for the probability scale, indicating the corresponding color values and their corresponding probability values.

We have also plotted a 'feature importance' bar chart. The feature importances help identify which features (in this case, 'votes' and 'cost for two') have a greater impact on the prediction. Higher importances indicates more influential features.

5] Relevance and Future Scope:

a) Relevance

This restaurant recommendation system is extremely relevant in the context of the food industry and consumer preferences. The technology gives significant insights and recommendations to restaurant owners and investors by leveraging machine learning techniques and data analysis. It addresses the following critical issues:

1. Decision-Making: The system supports users in making educated judgments on where to operate a new restaurant. Users can avoid risks and boost the possibilities of success for their restaurant initiatives by considering aspects such as votes, cost for two persons, and other characteristics.

2. Market Analysis: The exploratory data analysis and visualizations provided by the system lead to a better understanding of the restaurant business. It enables customers to acquire insights about the success likelihood of restaurants in various locations as well as uncover crucial elements driving their performance.

3. User Experience: The system provides an easy and engaging platform for users to submit their preferences and obtain customized recommendations by adding a user interface. It improves the user experience and allows for more efficient decision-making.

b) Future Scope

The restaurant recommendation system has potential for further development and expansion. Here are some avenues that can be explored to extend the project:

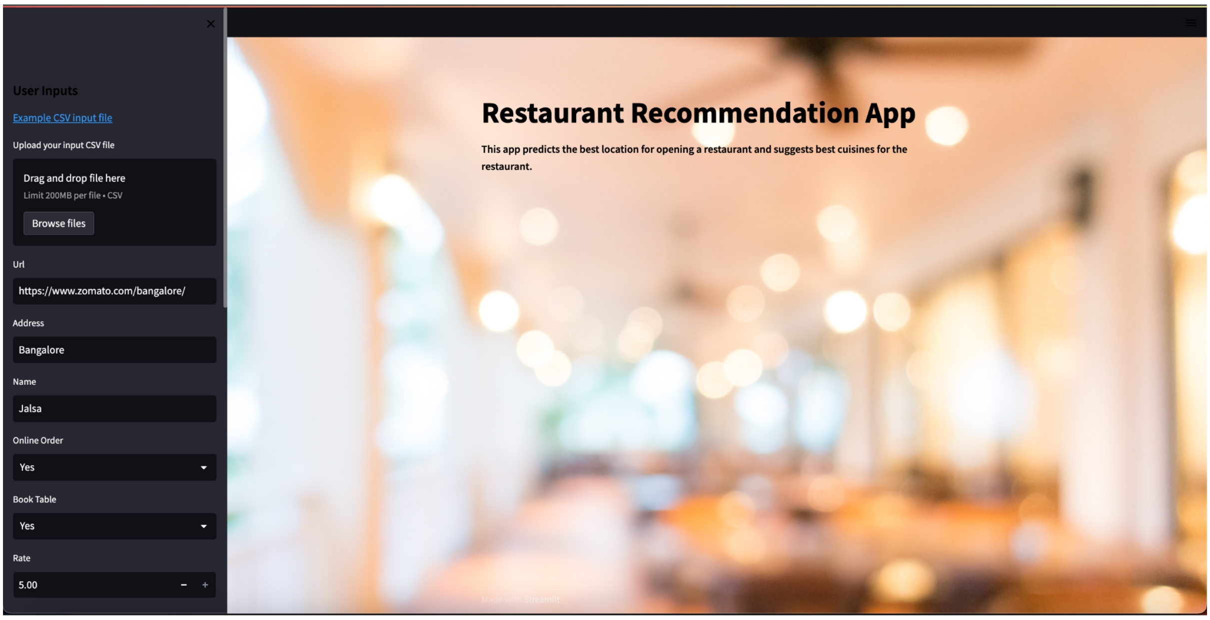
1. Advanced Recommendation Algorithms: While the current system uses the K-Nearest Neighbors (KNN) model, future improvements could include the incorporation of more advanced recommendation algorithms like collaborative filtering, content-based filtering, or hybrid models. Based on user choices and behavior, these algorithms can deliver more accurate and personalized recommendations.

2. Integration with Online Platforms: To collect real-time data on restaurant ratings, reviews, and customer feedback, the system can be coupled with online restaurant platforms and review aggregators. By combining this dynamic data, the system may provide consumers with more up-to-date and complete recommendations.

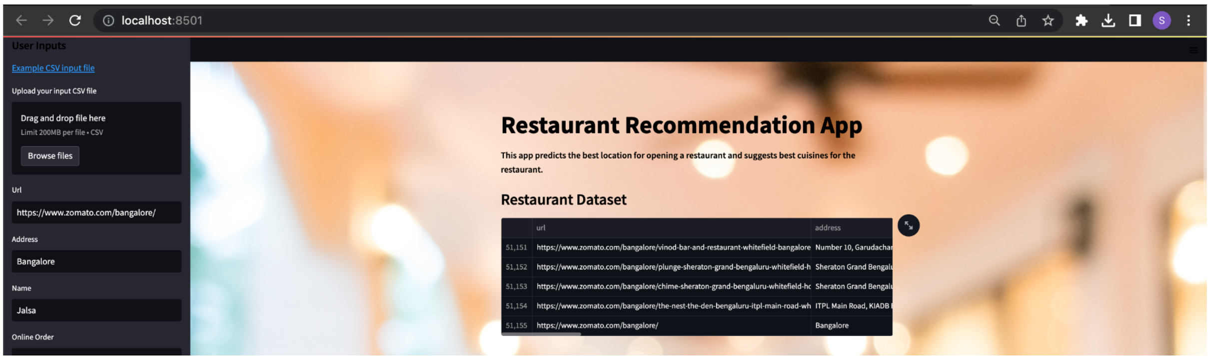
3. Incorporation of Sentiment Analysis: By incorporating sentiment analysis tools into the system, you can gain insights into client sentiments and views regarding eateries. The technology may uncover patterns and sentiment trends by evaluating customer reviews and feedback, allowing users to understand the reputation and customer happiness of various eateries.
4. Integration with Business Analytics: Users can undertake in-depth analyses of their own restaurant data by integrating the recommendation system with business analytics tools. Users can use data to make data-driven decisions and optimize their restaurant operations by analyzing sales patterns, customer demographics, and other business indicators.

In conclusion, the restaurant recommendation system provides significant insights and recommendations to restaurant owners and investors. Its value comes from assisting decision-making, offering market analysis, and improving the user experience. Furthermore, improved recommendation algorithms, integration with online platforms, sentiment analysis and integration with business analytics tools are all possibilities for future development.

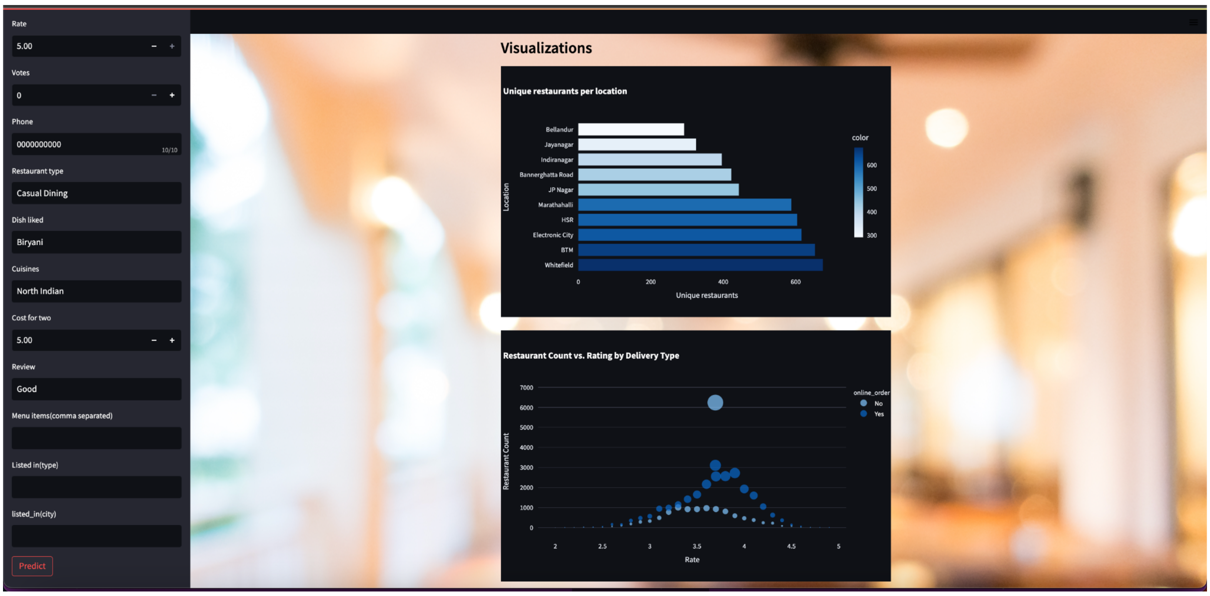
5] Demonstration Screenshots:



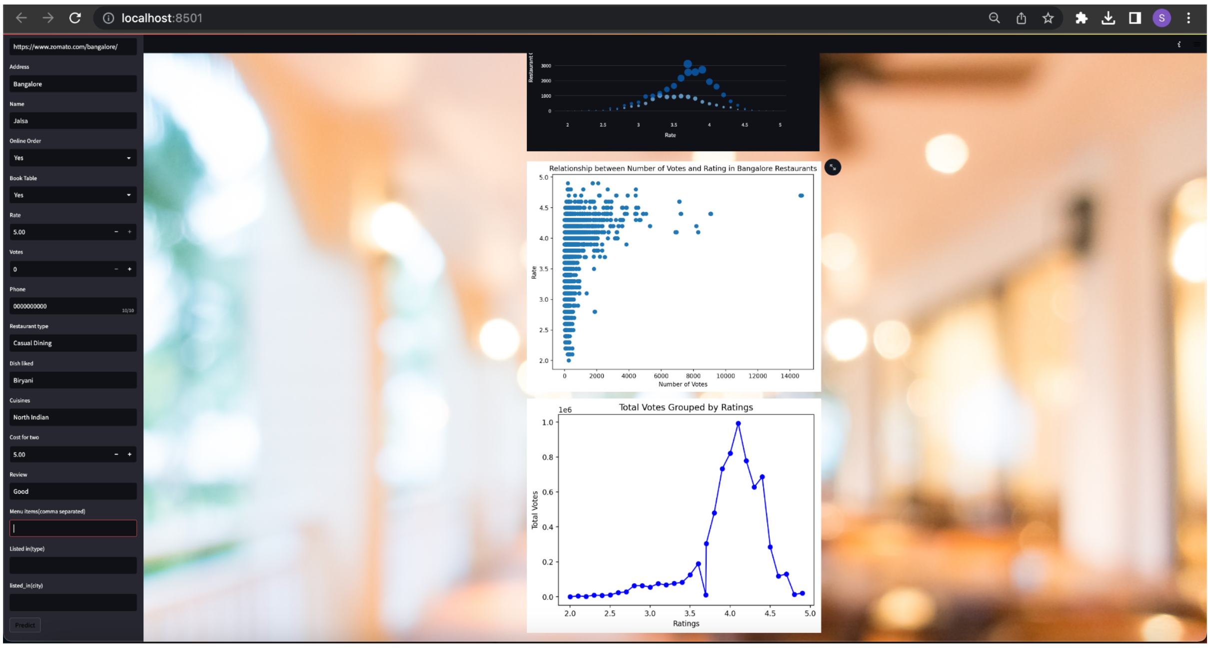
a) Front Display Page of our webpage



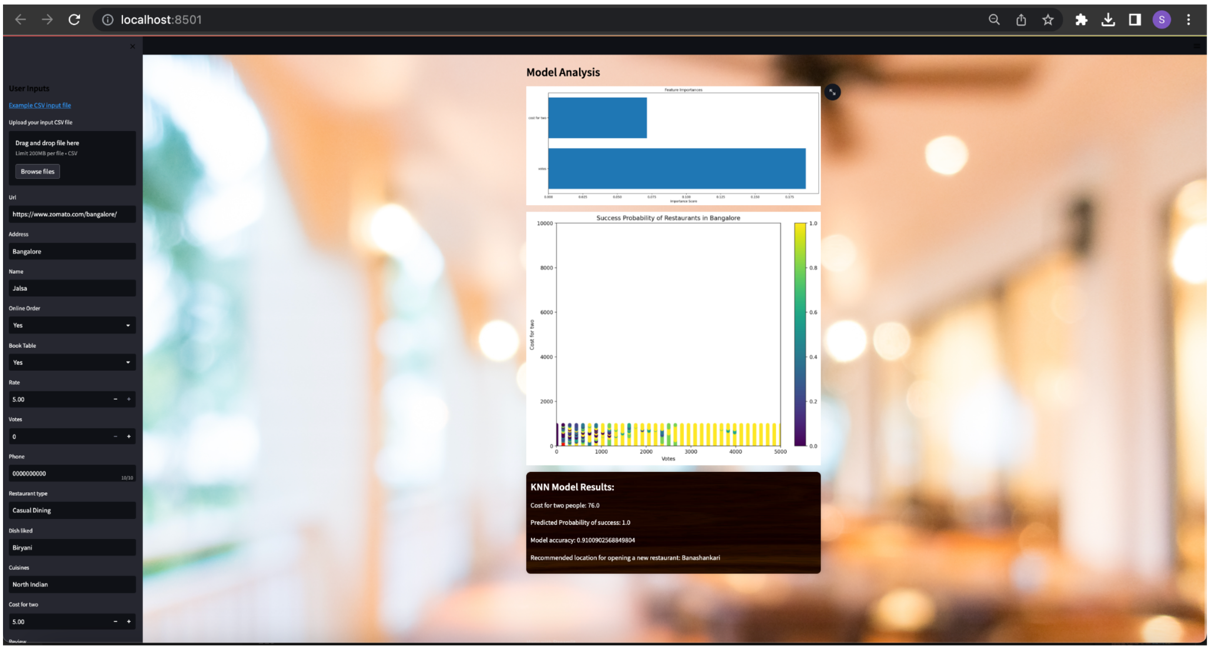
b) Loading the Dataset



c) Visualizations of our data



d) More visualizations



e) Model Analysis and Results

6] Output Analysis:

Based on the analysis performed by the restaurant recommendation system, the following recommendations and insights can be derived:

1. Location Selection: The system can provide recommendations for the best location to open a new restaurant based on various factors such as votes, cost for two people, and other relevant features. Users can leverage this information to make informed decisions regarding restaurant expansion or new ventures.

2. Understanding Success Factors: The system's analysis and visualizations shed light on the important features that contribute to the success of restaurants. Users can learn from these insights and focus on optimizing those factors to enhance

Investors and restaurant owners can use the exploratory data analysis completed in phase 1 and the models generated in phase 2 to analyze trends in restaurants and identify elements that may influence a restaurant's success.

As a result, the project's various phases' outcomes, as well as the project's future scope, are all highly relevant in the restaurant business.

7] <u>References:</u>

[1] https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants
[2] https://matplotlib.org/stable/plot_types/index
[3] https://pandas.pydata.org/docs/reference/general_functions.html
[4] https://seaborn.pydata.org/api.html
[5] https://www.w3schools.com/python/matplotlib_intro.asp
[6] https://www.techtarget.com/searchbusinessanalytics/definition
[7]https://towardsdatascience.com/all-machine-learning-models-explained-in-6-minutes-9fe30ff6776a
[8]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[9] https://machinelearningmastery.com/gentle-introduction-bag-words-model/
[10] https://realpython.com/build-recommendation-engine-collaborative-filtering/
[11] https://docs.streamlit.io/library/get-started