RestaurantRecommendation System

PreparedFor
Smart-Internz
AppliedDataScienceGuidedproject

By
Anuradha Bhimrao Kashid
DYPatilAgricultureandTechnicalUniversity,Talsande

On
23June2025

Abstract

This project develops a personalized restaurant recommendation system based on user preferences, location, and dining history. It analyzes factors such as cuisine, price, and ratings to suggest suitable diningoptions.Machinelearningtechniqueslikecollaborativeandcontent-basedfilteringareusedfor accurate suggestions. The system enhances the dining experience by offering relevant and location- aware recommendations

FinalProjectReport

Contents

1Introduction

1.1Projectoverviews

The **Restaurant Recommendation System** is a smart, data-driven solution designed to help usersefficientlydiscoverrestaurantsthatalignwiththeiruniquepreferencesandsituationalcontexts.As urbanization and mobile technologies continue to reshape consumer behavior, users are often overwhelmed by the sheer volume of available dining choices across platforms such as Google, Yelp, and Zomato. This leads to **decision fatigue** and suboptimal dining experiences. To solve this, the proposed system leverages a **hybrid recommendation model** combining collaborative filtering, content-based filtering, and geolocation-aware services.The **collaborative filtering component** analyzeshistoricaluserbehavior,includingpastrestaurantvisits,ratings,andinteractionpatterns,to identify users with similar tastes and recommend restaurants favored by like-minded individuals. Meanwhile,the**content-basedfilteringmodule**evaluatesrestaurantattributes—suchascuisinetype, price range, ambiance, and dietary offerings—to match them with explicit user preferences.
To enhance practicality, **geolocation data** is integrated using GPS APIs or IP-based location tracking. Thisallowsthesystemtodynamicallyadaptitsrecommendationsbasedontheuser'scurrentpositionor aspecifiedlocation,ensuringthatresultsareboth **relevantandaccessible**. Forexample,auserseeking budget-friendly vegan food in a new city would receive highly localized and personalized recommendations.

Furthermore, the system is designed with **adaptive learning capabilities**. Using techniques like reinforcementlearningorpreferencefeedbackloops,therecommendationengineimprovesovertimeby understanding user behavior patterns, modifying weightage of features, and incorporating real-time feedback such as likes, bookmarks, or direct reviews.

1.2 **Objectives**

1. **To design and implement a recommendation engine** that effectively filters and ranks restaurantsbasedonindividualuserpreferences,includingfoodtype,cost,ambiance,anddietary needs.
2. **Toapplymachinelearningmodels**,suchascollaborativefiltering(user-basedanditem-based) and content-based filtering, to identify patterns in user behavior and restaurant attributes.
3. **Toincorporatelocation-awarefeatures**usingGPSoruser-inputtedlocationdata,ensuringthat recommended restaurants are conveniently accessible to the user.
4. **Togatherandanalyzerestaurantreviewsandratings** frompublicsources(e.g.,Yelp,Google Reviews, or internal datasets) to improve the trustworthiness and relevance of suggestions.
5. **Tocreateauser-friendlyinterface**thatallowsuserstoinputpreferences,viewrecommended restaurants, and interact with the system seamlessly.
6. **Todevelopafeedbackmechanism** thatcollectsusersatisfactiondatapost-visittorefinefuture recommendations and enhance personalization over time.
7. **Toensurescalabilityandadaptability**ofthesystemforuseindifferentgeographicregionsor for integration into existing food delivery or travel applications.

SMARTBRIDGE Let's Bridge the Gap a Veranda Enterprise

Smart Internz

2ProjectInitializationandPlanningPhase

2.1DefineProblemStatement

2ProjectInitializationandPlanningPhase

ProblemStatements(RestaurantRecommendationsystem):

PS No.	I am (Customer)	I'mtryingto	But	Because	Whichmakesme feel
PS-1	Atouristina new city	Findgoodlocal restaurants	Idon'tknowthe area well	I lack local knowledge or reviews	Confused and unsureofwhereto eat
PS-2	Avegetarian diner	Get recommendations for veg-only restaurants	Most apps showmixed cuisineplaces	I want strict dietaryoptions	Frustratedand unsupported
PS-3	Arestaurant owner	Attractmore customers through recommendation platforms	Myrestaurantis not being recommended often	The system doesn't promote new or small businesses	Invisibleand discouraged
PS-4	A student on a tight budget	Findaffordable but tasty restaurants	Expensive optionsare shownfirst	Filters don't prioritizepriceor value	Overwhelmedand discouraged
PS-5	Adeliveryapp user	Getsuggestions based on past orders	It doesn't adapt to my taste	Thesystemlacks learning	Frustratedby repetition
PS-6	Aparentof youngkids	Findkid-friendly and hygienic restaurants	No way to filter forchild-friendly	Lackofsafetyand family-focused featuresamenities	Anxiousabout experience

PS-7	A small restaurant owner	Increase customer footfallvia platforms	Mybusinessis buriedunder chainlistings	Ranking algorithms favorlarge brands	Discouragedand invisibleguide
PS-8	Anew-in-town resident	Explore culturally diversefood options	Unaware of hiddengemsin my area	No cultural/ethnic tagsoruserreviews	Disconnectedand bored of same cuisine
PS-9	Afooddelivery platformanalyst	Monitorfood safety and restaurant quality	Can'tverify ingredientsafety from menus	Platforms lack AI food item scanners or trackers	Concernedabout consumer trust
PS-10	Adatascientist	Analyzefood trends from reviews	Datasets are messy,biased,or unavailable	Lack of structured sentimen and metadata	Blockedinmodel building and research
PS-11	A foodie traveler	Findtop-rated local restaurants in new cities	Recommendations don'tmatchmy tasteor location	Generic,irrelevant suggestions	Frustrated and unsurewhereto eat
PS-12	Arestaurant owner	Improvemy visibilityon food apps	My reviews are outdatedorlow-rated	I can't easily respondorupdate info	Powerlessand misrepresented
PS-13	A health-conscious customer	Find healthy eating options nearby	Menusandcalorie info are missing	I can't make informeddecisions	Disconnected frommyhealth goals
PS-14	**A** healthconscious individual	Track the health benefits of different mushrooms	I can't identifywhat'sin the store or dish	There'snoeasyapp forinstant scanning	Disappointed and disconnected frommyhealth goals

ProjectProposal(ProposedSolution)

ProjectProposal(Proposed Solution)

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach,keyfeatures,andresourcerequirements,includinghardware,software,andpersonnel

ProjectOverview	
Objective	Todevelopasystemthatprovidespersonalizedandefficientrestaurant recommendationsby analyzing user preferences,dietary requirements, location, and budget.
Scope	The project aims to serve users seeking restaurant suggestions that match their individual lifestyle choices and dining preferences. It will operate acrossvariousregions,consideringreal-timedataandqualitativereviews.
ProblemStatement	
Description	Finding restaurants tailored to specific needs is often time-consuming and inefficient.Usersfrequentlyrevisitthesameplaces,missingdiverseoptions that better match their preferences.
Impact	Solvingthisproblemimprovesusersatisfaction,encouragesexplorationof new dining options, and reduces time spent on decision-making.
ProposedSolution	
Approach	Thesolutionemploysinnovativerecommendationalgorithmsthatfactorin both user input and external data like ambiance, ratings, and reviews. It adapts dynamically to user feedback and real-time changes.
Key Features	• Personalizedrecommendations • Real-timedataanalysis • Integrationofuserreviews • Considerationofdietaryandbudget constraints • Scalableinfrastructure

ResourceRequirements

ResourceType	Description	Specification/Allocation
Hardware		
ComputingResources	8-coreCPUsandoptional GPU	2xNVIDIAV100GPUs
Memory	RAM	Minimum8GBRAM
Storage	SSD	1TBSSDforstoringuserdata and restaurant metadata
Software		
Frameworks	Python frameworks	Python, Flask
Libraries	Additionallibraries	Pandas, NumPy, Scikit-learn, TensorFlow,BeautifulSoup(for scraping), and NLTK (for review analysis)
DevelopmentEnvironment	IDE,versioncontrol	Jupyter Notebook
Data		
Data	Size:-Approx.50,000–100,000 recordsinitially;scalablebased on user growth, Format:-CSVfortabular datasets, Text/HTML for scraped reviews	Aggregatedfromcrowdsourced restaurantplatforms(e.g.,Yelp, Zomato APIs), user feedback, and public review datasets

InitialProjectPlanning

ProductBacklog,SprintSchedule,andEstimation

Sprint	Functional Requirement (Epic)	User Story Number	UserStory/ Task	Story Points	Priority	Sprint StartDate	Sprint EndDate (Planned)
Sprint-1	User Preferences Input	USN-1	As a user, I can enter my foodor Hotel preferences.	2	High	01 June 2025	02 June 2025
Sprint-1	Recommendation Engine	USN-2	As a user, I can get restaurant recommendations based on my preferences.	3	High	02 June 2025	02 June 2025
Sprint-2	Review &Rating Integration	USN-3	As a user, I can view restaurant reviews and ratings fetchedfrom the dataset.	2	Medium	03 June 2025	04 June 2025
Sprint-2	UI/UX Enhancement	USN-4	As a user, I can view results in a user-friendly interfacewith filters and sorting.	2	Medium	04 June 2025	05 June 2025

2 DataCollectionandPreprocessingPhase

DataCollectionPlanandRawDataSourcesIdentified

DataCollection Plan

Section	Description
Project Overview	Developarestaurantrecommendationsystemtoassistusersinfinding dining options based on their preferences, location, and other relevant factors.Byanalyzinguserpreferences,restaurantratings,andlocation data, this project aims to provide personalized recommendations that enhance the dining experience for users.
DataCollection Plan	The dataset used for this project was sourced from Kaggle and contains detailed information on over 9,000 restaurants in Bangalore, including attributeslikename,location,cuisine,ratings,andpricing.Thispublicly available dataset was collected to support analysis and predictive modeling related to restaurant ratings and customer preferences.
RawDataSources Identified	Therawdataforthisproject wasobtainedfromtheKaggledataset titled **"ZomatoBangaloreRestaurants"**byHimanshuPoddar.Thedatasetis publicly available at https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurantsandincludeskeyrestaurant-relatedattributessuchasrestaurant names, locations, cuisines, average costs, online delivery availability, and user ratings.

RawDataSources

Source Name	Description	Location/URL	Format	Size	Access Permissions
SmartInterz Provided Dataset	Restaurant-leveldata including name, location, cuisines, rating and cost.	Data-Set zomato-bangalore-restaurants	CSV	~ 93MB	Public

2.2DataQualityReport

DataSource	DataQualityIssue	Severity	ResolutionPlan
Dataset (Restaurant reviewsand metadata)	Missingvaluesinfields like restaurant name, location, or ratings	Moderate	Performdataimputationusingtechniques like mean/mode for numeric values and most frequent value for categorical data. Alternatively, remove rows with critical missing fields.
Dataset(User reviews)	Duplicateuserreview entries	Low	Remove duplicate records using drop_duplicates() in pandas or SQL DISTINCTqueries.Usedatetimeparsing libraries (e.g., pandas.to_datet ime) to standardize all date/time fields.
Dataset (Restaurant metadata)	Inconsistent formats (e.g.,locationwrittenin different ways like "NY", "New York")	Moderate	Apply data standardization techniques, usingstringfunctionsorregexpatternsto unify the format.
Dataset(User preferenc es)	Sparse data or insufficientuserhistory	High	Implement fallback strategies such as popularity-based or content-based recommendationswhenuserdataislacking.

2.3DataPreprocessing

DataPreprocessing

The images will be preprocessed by resizing, normalizing, augmenting, denoising, adjusting contrast,detectingedges,convertingcolorspace,cropping,batchnormalizing,andwhitening data. These steps will enhance data quality, promote model generalization, and improve convergenceduringneuralnetworktraining,ensuringrobustandefficientperformanceacross various computer vision tasks.

Section	**Description**
Data Overview	The dataset contains restaurant information from Zomato, includingname,reviews,ratings,cuisines,cost,andmore.The data is cleaned, deduplicated, and preprocessed for building a content-based recommendation system.
Resizing	*Notapplicablefortext data.*
Normalization	Ratingsarenormalizedtoa1-5scaleusingMinMaxScaler. Text is lowercased and punctuation is removed.
Data Augmentation	Notapplicablefortextdata.
Denoising	Textiscleanedbyremovingnewlinecharactersand punctuation.
Edge Detection	Notapplicablefortextdata.
ColorSpace Conversion	Notapplicablefortextdata.
ImageCropping	Notapplicablefortextdata.
BatchNormalization	Notapplicablefortextdata.
DataPreprocessingCode Screenshots	

<table>
<tr>
<td>LoadingData</td>
<td>
<pre><code># Mounting Google Drive
#from google.colab import drive
#drive.mount('/content/drive')
import csv
# Specifying the path to the dataset file
file_path = '/content/zomato.csv'

Reading the dataset into a Pandas DataFrame
#df = pd.read_csv(file_path,encoding = 'ISO-8859-1', low_memory = False)
df = pd.read_csv(file_path, encoding='ISO-8859-1', on_bad_lines='skip', engine='python')

Displaying the first few rows of the dataset to ensure it's loaded correctly
df.head()</code></pre>
Python
</td>
</tr>
<tr>
<td>Resizing</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Normalization</td>
<td>
<pre><code># Computing Mean Rating
restaurants = list(df['name'].unique())
df['Mean Rating'] = 0
for i in range(len(restaurants)):
    df['Mean Rating'][df['name'] == restaurants[i]] = df['rate'][df['name'] == restaurants[i]].mean()
#Scaling the mean rating values
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler (feature_range = (1,5))
df[['Mean Rating']] = scaler.fit_transform(df[['Mean Rating']]).round(2)</code></pre>
</td>
</tr>
<tr>
<td>Data Augmentation</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Denoising</td>
<td>
<pre><code>## Lower Casing
df["reviews_list"] = df["reviews_list"].str.lower()
Removal of Puctuations
import string
PUNCT_TO_REMOVE = string.punctuation
def remove_punctuation(text):
    """custom function to remove the punctuation"""
    return text.translate(str.maketrans('', '', PUNCT_TO_REMOVE))
df["reviews_list"] = df["reviews_list"].apply(lambda text: remove_punctuation (text))</code></pre>
</td>
</tr>
<tr>
<td>Edge Detection</td>
<td>Not applicable</td>
</tr>
<tr>
<td>ColorSpace Conversion</td>
<td>Not applicable</td>
</tr>
<tr>
<td>ImageCropping</td>
<td>Not applicable</td>
</tr>
<tr>
<td>BatchNormalization</td>
<td>Not applicable</td>
</tr>
</table>

4.ModelDevelopmentPhase

2.4ModelSelectionReport

Model	Description
Content-Based Filtering	Content-basedfilteringrecommendsrestaurantsbycomparinguserpreferences (e.g.,cuisinetype,pricerange,dietaryrestrictions)withrestaurantattributes.It focuses on similarities between items and the user's profile without relying on other users' data. Thismethod iseffective for users with unique tastes butmay strugglewithlimiteduserprofiles(coldstart).
Collaborative Filtering	Collaborative filtering leverages the preferences of similar users to make recommendations. Ituses historical ratingsandreviews toidentify patterns. This modeliseffectiveindiscoveringnewitemsbutcansufferfromsparsityandcold startproblemsifdataislimited.
Hybrid Recommendation Model	This combines content-based and collaborative filtering to overcome the limitationsofeachmethod.Byintegratingbothuserpreferencedataandbehavior ofsimilarusers,hybridmodelsimproverecommendationaccuracy,diversity,and scalability. It is particularly useful in scenarios with large, sparse datasets like restaurantrecommendations.
Matrix Factorization	Matrix factorization techniques decompose the user-item interaction matrix into latentfeatures,capturingunderlyingpatternsin user preferences. SingularValue Decomposition(SVD)isacommonapproach.Itiscomputationallyefficientand workswellforlargedatasetsbutrequiresenoughratings.
DeepLearning (Neural Networks)	Neural networks can be used to build recommendation systems by learning complex,non-linearrelationshipsbetweenusersandrestaurantsfromrichfeature sets including reviews, preferences, and metadata. While powerful, they require large datasets and are computationally intensive.

Conclusion:

ModelSelected	
Hybrid Recommendation Model	The hybrid model was selected because it addresses the limitations of both content-basedandcollaborativefilteringapproaches.Iteffectivelyhandlesthecoldstartand sparsity issues by integrating multiple data sources such as user profiles, restaurant attributes, and behavioral data. This results in more personalized, diverse, and accurate recommendations, making it highly suitable for a restaurant recommendationsystemwithvaryinguserpreferencesanddataavailability.

2.5InitialModelTrainingCode,ModelValidationandEvaluation Report

InitialModelTrainingCode,ModelValidationandEvaluationReport Initial Model Training Code (5 marks):

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
df_percent.set_index('name', inplace=True)
indices = pd.Series(df_percent.index)

# Creating tf-idf matrix
tfidf = TfidfVectorizer(analyzer='word', ngram_range=(1, 2), min_df=0, sto
tfidf_matrix = tfidf.fit_transform(df_percent['reviews_list'])

cosine_similarities = linear_kernel(tfidf_matrix, tfidf_matrix)
```

ModelValidationandEvaluationReport(5marks):

Model	Summary	TrainingandValidationPerformance Metrics
Model1	Content-basedRecommendation	**TrainingMetrics-**None(unsupervised,no explicit training phase) **Validation Metrics - **None (recommendationsareinspectedmanually)

3ModelOptimizationandTuning Phase

3.1TunningDocumentation

HyperparameterTuning

<table>
<tr><th>Model</th><th>TunedHyperparameters</th></tr>
<tr><td>Model 1:
Content-Based
Filtering</td><td>- SimilarityMetric:Cosinesimilaritywasusedastheprimarymetricto compute similarity between restaurants based on features like cuisines, rating, and cost.
- TopNRecommendations:Thenumberoftopsimilarrestaurants returned was tested with values like 5, 10, and 15.
<pre><code>def recommend(name, cosine_similarities = cosine_similarities):

    # Create a list to put top restaurants
    recommend_restaurant = []

    # Find the index of the hotel entered
    idx = indices[indices == name].index[0]

    # Find the restaurants with a similar cosine-sim value and order them from bigges number
    score_series = pd.Series(cosine_similarities[idx]).sort_values(ascending=False)

    # Extract top 30 restaurant indexes with a similar cosine-sim value
    top30_indexes = list(score_series.iloc[0:31].index)

    # Names of the top 30 restaurants
    for each in top30_indexes:
        recommend_restaurant.append(list(df_percent.index)[each])

    # Creating the new data set to show similar restaurants
    df_new = pd.DataFrame(columns=['cuisines', 'Mean Rating', 'cost'])

    # Create the top 30 similar restaurants with some of their columns
    for each in recommend_restaurant:
        df_new = df_new.append(pd.DataFrame(df_percent[['cuisines','Mean Rating', 'cost']][df_percent.index == each].sample()))

    # Drop the same named restaurants and sort only the top 10 by the highest rating
    df_new = df_new.drop_duplicates(subset=['cuisines','Mean Rating', 'cost'], keep=False)
    df_new = df_new.sort_values(by='Mean Rating', ascending=False).head(10)

    print('TOP %s RESTAURANTS LIKE %s WITH SIMILAR REVIEWS: ' % (str(len(df_new)), name))
    df_new.index = df_new.index.str.lower()
    return df_new</code></pre></td></tr>
</table>

Model 2: Collaborative Filtering	- **Algorithm:**SVD(SingularValueDecomposition)fromtheSurprise library. - **LearningRate:**Tuned valuessuchas0.005,0.01,and0.02weretested. - **Regularization:**Parameterssuchas0.02,0.05weretriedtoavoid overfitting. - **NumberofEpochs:** Adjustedbetween20 and100 epochs.

FinalModelSelectionJustification

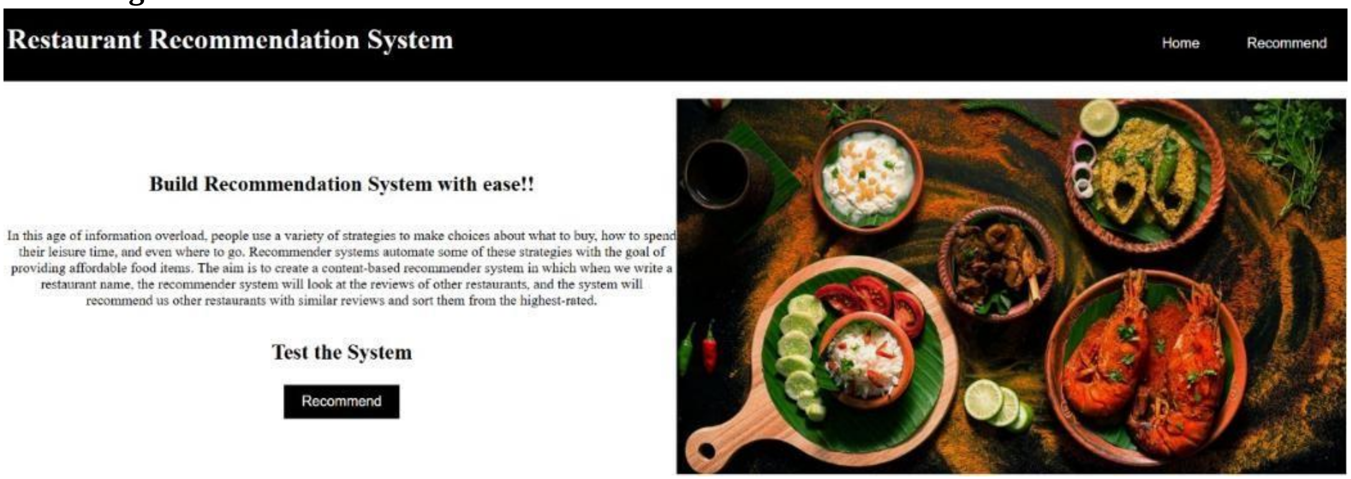
FinalModelSelection Justification:

Final Model	**Reasoning**
Model1:Content-Based Filtering	Selectedduetoitssimplicityandgoodperformancewithoutrequiring detailed user history. It gave interpretable and relevant results using restaurant features like cuisines, ratings, and cost.

4Results

4.1OutputScreenshots

Home Page:



InputPage:

Example:-

Restaurant Recommendation System Home Recommend

Here are the top recommended restaurants

Name	Cuisines	Mean Rating (out of 5)	Cost (in thousands)
The Black Pearl	north indian european mediterranean bbq	4.85	1.5
Barbeque Nation	north indian european mediterranean bbq kebab	4.7	1.6
Hunger Camp	north indian south indian chinese seafood	4.56	1.3
Hakuna Matata	north indian asian seafood chinese	4.41	1.2
Jalsa Gold	north indian mughlai italian	4.41	1.3
Deja Vu Resto Bar	north indian italian	4.26	900.0
Tipsy Bull - The Bar Exchange	north indian chinese continental mexican	4.26	1.4
Dhaba Estd 1986 Delhi	north indian	4.26	1.1
Float	north indian japanese	4.26	1.5
nu.tree	north indian healthy food beverages	4.26	400.0

5Advantages& Disadvantages

Advantages:

- **PersonalizedUserExperience**:Tailorsdiningoptionsbasedonuserpreferences,dietaryneeds, and previous behaviour.
- **Time-saving**:Reducestheeffortneededtosearchandchoosea restaurant.
- **ImprovedDiscoverability**:Helpssmallerornewrestaurantsgainvisibilitythrough recommendations.
- **Data-DrivenDecisions**:Usesuserratings,reviews,andlocationdatatomakeinformed suggestions.
- **EnhancedCustomerSatisfaction**:Usersaremorelikelytoenjoytheirmealswhen recommendations align with their preferences

Disadvantages:

- **PrivacyConcerns**:Collectingandanalyzinguserdata(location,preferences)canraiseprivacy issues.
- **BiasinRecommendations**:Algorithmsmightfavorsponsoredlistingsorhigh-traffic restaurants, reducing diversity.
- **DependenceonUserData**:Inaccurateorlimiteddatacanleadtopoorrecommendations.
- **Over-Personalization**:Usersmightbeconfinedtosimilarchoices,missingoutonnewor diverse dining experiences.
- **ScalabilityIssues**:Maintainingsystemaccuracyandperformancecanbecomechallengingas the user base grows.

6Conclusion

A restaurant recommendation system is a powerful tool for enhancing the dining experience by delivering tailored suggestions based on user behavior, preferences, and location. While it offers significant benefits such as convenience, personalization, and efficient decision-making, it also presentschallengesincludingdataprivacy,systembias,andtheriskofuserdatadependency.Future advancements in AI, real-time analytics, and user interface technologies promise to make such systems more intelligent, inclusive, and immersive. With careful implementation and ethical considerations, this system can transform how users explore and enjoy culinary option

7FutureScope

- **IntegrationwithAR/VR**:Inthefuture,userscouldtakevirtualtoursofrestaurantsorview their ambiance in AR before booking.
- **VoiceAssistantCompatibility**:IntegrationwithSiri,Alexa,orGoogleAssistanttoprovide hands-free restaurant suggestions.
- **EnhancedPersonalization**:Usedeeplearningandbehavioralanalyticstorefine suggestions based on dietary restrictions, allergies, and eating habits.
- **Real-timeDataUtilization**:Incorporatingreal-timefactorslikewaittimes,specialoffers, and crowd density for more dynamic recommendations.
- **MultilingualSupport**:Expandingthesystemtosupportvariouslanguagestocatertoa global audience.
- **SocialMediaIntegration**:Useofsocialmediatrendsandcheck-instoimprove recommendation relevance.
- **SustainabilityPreferences**:Factoringineco-consciousdiningchoices(e.g.,locally sourced, plant-based, or low-waste restaurants).

8 Appendix

8.1SourceCode

[Restaurant-Recommendation-System]

8.2ProjectVideoDemoLink :

VideoDemoLink:
[Demo Link]