The IOT Academy

Capstone Project – 02

Solution Design Document v1.0

Zomato Recommendation

Date : 15 - 03 - 2024

[Introduction 3](#_Toc161416703)

[Project Team 4](#_Toc161416704)

[Methodology 5](#_Toc161416705)

[Breakdown of the notebook: 5](#_Toc161416706)

[github.com URL 5](#_Toc161416707)

[Load data and import 5](#_Toc161416708)

[EDA 5](#_Toc161416709)

[Text Preprocessing 6](#_Toc161416710)

[Recommendation 6](#_Toc161416711)

[conclusion 7](#_Toc161416712)

# Introduction

In today's fast-paced world, dining out has become an integral part of urban lifestyle. With an abundance of restaurants offering diverse cuisines, selecting the perfect dining spot can often be a daunting task for consumers. To address this challenge, the integration of Natural Language Processing (NLP) techniques into restaurant recommendation systems has gained significant traction. Zomato, one of the leading online food delivery and restaurant discovery platforms, stands at the forefront of leveraging NLP to enhance user experience and satisfaction.

The aim of this project is to develop a robust and accurate restaurant recommendation system using NLP techniques applied to Zomato's vast dataset. By harnessing the power of NLP, we seek to provide users with personalized restaurant recommendations tailored to their preferences, dietary restrictions, location, and other relevant factors.

Through this project, we endeavor to explore various NLP methodologies, including sentiment analysis, topic modeling, and collaborative filtering, to extract meaningful insights from user reviews, menu descriptions, and other textual data available on the Zomato platform. By analyzing and understanding the implicit and explicit user feedback embedded within these textual sources, we aim to build a recommendation engine that not only suggests restaurants based on user preferences but also delivers a delightful and personalized dining experience.

Furthermore, this project will delve into the challenges and opportunities associated with processing and analyzing large-scale textual data from diverse sources such as user reviews, restaurant descriptions, and menu items. We will employ state-of-the-art NLP algorithms and techniques to preprocess, analyze, and extract valuable information from unstructured text data, thereby enabling accurate and effective restaurant recommendations.

In summary, this project represents an exciting endeavor to harness the potential of NLP in revolutionizing the restaurant recommendation landscape. By leveraging advanced NLP techniques and Zomato's extensive dataset, we aspire to develop a sophisticated recommendation system that enhances user satisfaction, fosters culinary exploration, and promotes the discovery of delightful dining experiences.

# Project Team

|  |  |  |
| --- | --- | --- |
| Name | Designation | Organization |
| Bhupendra Prasad | Senior Solution Architect | Atos (Dubai,U.A.E) |
| Yash Gangurde | Senior Technical Consultant | MATRIXX Software(Dubai,U.A.E) |

# Methodology

Recommendation Systems help improve search results by suggesting items that are relevant to users' interests or related to their search history. They personalize information based on user preferences, making them active information filtering systems.

There are three main types of recommendation systems:

1. **Demographic Filtering**: Offers generalized recommendations based on popularity or genre. It suggests the same items to users with similar demographic features.
2. **Content-Based Filtering**: Recommends similar items based on specific attributes of a chosen item. This system uses item metadata like genre, director, or description to make recommendations.
3. **Collaborative Filtering**: Matches users with similar interests and suggests items based on this match. It doesn't require item metadata.

In this project, we focus on Content-Based Filtering. This method uses information about the description and attributes of items previously consumed by the user to model preferences. It recommends items similar to those previously liked or examined by the user. The system compares candidate items with previously rated items by the user and recommends the best-matching ones.

The dataset used contains restaurants from Bangalore, India, collected from Zomato.

**Objective**: To create a content-based recommender system where, upon inputting a restaurant name, the system analyses reviews from other restaurants. It then recommends similar restaurants with high ratings, sorted by relevance.

## Breakdown of the notebook:

#### github.com URL

Please check below URL for code:

|  |
| --- |
| https://github.com/Bhupendra-Prasad/The-IoT-Academy/blob/Projects/Zomato\_Recommendation.ipynb |

#### Load data and import

Load the data from zomato.csv and import below libraries.

|  |
| --- |
| import numpy as np  import pandas as pd  import seaborn as sns  import matplotlib.pyplot as plt  import warnings  warnings.filterwarnings("ignore")  import re |

#### EDA

We have done following steps:

* Deleting redundant columns.
* Dropping duplicates.
* Cleaning individual columns.
* Remove the NaN values from the dataset.
* Check if same restaurant occurs multiple times.
* Check where rate and reviews list both are null
* Removing rows with rate as - and NEW from data frame.
* Removing slash and space from rate.
* TypeCasting rate column into float from object.
* Getting Mean Rating of the Restaurants for analysis.
* Merging MeanRating of restaurants with datafarame
* Dropping rows where mean rating and rate are nulls.

### Text Preprocessing

Following steps, we have performed to prepare the test:

* Cleaning unnecessary words in the reviews
* Lower casing
* Removal of Punctuations
* Removal of Stopwords
* Removal of URLs
* Spelling correction
* Removing links and other unnecessary items
* Removing Symbols
* Removing space from rate
* Removing space from rate
* Lower Case conversion reviews\_list inplace
* Basic Cleaning remove e.g: URL,ctrl/ASCII char,spl char def cleanText(Zomtext)
* Calling function cleanText
* Remove char,spl char
* removal of stopwords present in stopwordList
* Calculating words frequency using “FreqDist”
* Creating Wordcloud for different ratings
* Creating tf-idf matrix
* Setting name as index
* Mapping of indices and restaurant names
* Initializing TF-IDF
* Compute the linear kernel

### Recommendation

Following steps are performed:

* Create a list recommend\_restaurant
* Find the index of the hotel entered
* Find the restaurants with a similar cosine-sim value and order them from bigges number
* Extract top 30 restaurant indexes with a similar cosine-sim value
* Find Names of the top 30 restaurants
* Creating the new data set to show similar restaurants
* Create the top 30 similar restaurants with some of their columns
* Drop the same named restaurants and sort only the top 10 by the highest rating

### conclusion

After analysing reviews from other restaurants on Zomato, our content-based recommender system has identified the top 10 restaurants that are similar to Eggzotic based on their reviews. These recommendations are sorted by relevance and are expected to provide Zomato.

