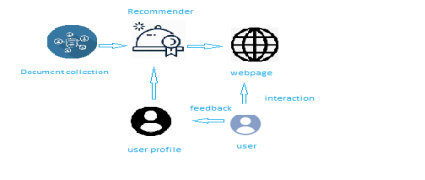
**PROJECT DESCRIPTION**

As we are users of recommendation applications, people care more about how we will like a restaurant. It is very common that we hang out with families, friends, and co-workers. when comes to lunch or dinner time. In the past, people obtained suggestions for restaurants from friends. Although this method is straightforward and user-friendly, it has some severe limitations. First, the recommendations from friends or other common people are limited to those places they have visited before. Thus, the user is not able to gain information about places less visited by their friends. Besides that, there is a chance of users not liking the place recommended by their friends.

**SOLUTION**

Here we are creating a content-based recommendation system. The aim is to create a content-based recommender system in which when we will write a restaurant name, the Recommender system will **look at the reviews of other restaurants**, and the System will recommend us other restaurants with **similar reviews** and sort them from the **highest-rated**. The main people who are going to benefit from this recommendation system are the tourists, who are new to a city. Most of the tourists always love to visit famous restaurants in a particular city during their visit. Otherwise, it can be heavily used by people belonging to the same city, to see if any new restaurant is recommended based on their activity.

**ARCHITECTURE**

****

**LEARNING OUTCOMES**

By the end of this project:

● You’ll be able to perform one of the techniques to build you recommendation system

● You’ll be able to know the recommendation system using Content-Based Filtering.

● You will be able to know how to pre-process / clean the data using different data pre-processing techniques.

● You will able to analyse or get insights of data through visualization. ● Applying algorithms according to dataset and based on visualization. ● You will able to know how to find accuracy of the model.

● You will be able to know how to build a web application using Flask framework.

**PRE-REQUISITES:**

To complete the project successfully, you need to install following software & packages:

**Activity 1: Install Anaconda IDE / Anaconda Navigator.**

**●** In order develop a solution to this problem statement, we need an environment to write and test the code.

**●** We use Anaconda IDE (Integrated Developing Environment). ● Refer to the below link to download & install Anaconda Navigator.

**Link:** https://www.youtube.com/watch?v=5mDYijMfSzs

**1. Activity 2: To build Machine learning models you must require the following packages**

● **Numpy**:

It is an open-source numerical Python library. It contains a multidimensional array and matrix data structures and can be used to perform mathematical operations

● **Numpy**:

It is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and

k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy

● **Matplotlib and Seaborn**

Matplotlib is mainly deployed for basic plotting. Visualization

using Matplotlib generally consists of bars, pies, lines, scatter plots and so on. Seaborn: Seaborn, on the other hand, provides a variety of visualization patterns. It uses fewer syntax and has easily interesting default themes.

**● Flask:**

Web framework used for building Web applications

If you are using **anaconda navigator**, follow below steps to download required packages:

● Open anaconda prompt.

● Type “pip install pandas” and click enter.

● Type “pip install matplotlib” and click enter.

● Type “pip install seaborn” and click enter.

● Type “pip install plotly” and click enter.

● Type “pip install numpy” and click enter.

● Type “pip install scikit-image” and click enter.

● Type “pip install scikit-learn” and click enter.

● Type “pip install Flask” and click enter.

**Link:** Introduction to Scikit-Learn (sklearn) in Python • datagy

**PRIOR KNOWLEDGE**

One should have knowledge on the following Concepts:

**Link:** Supervised and Unsupervised Learning

Watch the below video to know about the types of machine learning **Link:** Regression, Classification and Clustering

**Link:** ML - Content Based Recommender System - GeeksforGeeks **Link:** NLTK :: Natural Language Toolkit

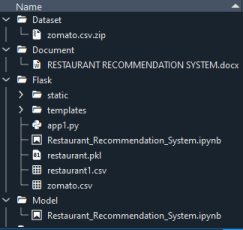
**Link:** Flask:

**Link:** Recommendation System

It is recommended to watch above video’s to understand the concepts before you start your project.

**PROJECT WORK FLOW**

● User interacts with the UI (User Interface) to enter the input features. ● Entered features/input is analysed by the model which is integrated ● Once model analyses the entered inputs, the prediction is showcased on the UI.



**TASKS**

1. Data Collection.

● Collect the dataset or Create the dataset

2. Data Pre- processing.

Import the Libraries.

Importing the dataset.

Exploratory Data Analysis

Data Visualization.

3. Content Based Filtering

Merging datasets

Creating the recommender system

Predicting the results

4. Application Building

Create an HTML file

Build a Python Code

**Milestone 1: Data Collection**

Now, the milestone-2 is all about creation or collection of dataset.

we will use the **Zomato** Bangalore for our analysis to draw conclusions using the content filtering method.

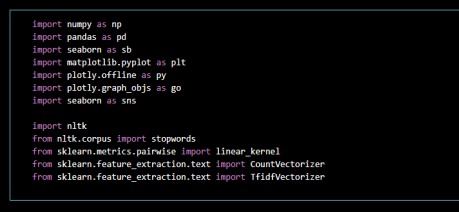
Here is the dataset link: **dataset**

**Milestone 2: Data Pre-processing**

In this milestone, you need to complete all the below activities to build the model.

**Activity 1: Import Libraries**

Import the below essential libraries for data pre-processing and creating recommendation system. Pandas and NumPy are used for data pre-processing and cleaning. Seaborn, Plotly and Matplotlib helped in creating visual graphics and bar plots for the dataset. Also, since there would be cleaning of text data (reviews) as well, therefore for that we will use nltk and sklearn library.

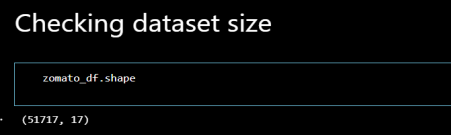


**Activity 2: Read the Dataset:**

Our dataset format might be in .csv, excel files, .txt, json, etc. We can read the dataset with the help of pandas.

In pandas we have a function called read\_csv () to read the dataset. As a parameter we have to give the directory of csv file.

**Activity 3: Analyse the Dataset:**

****The dataset contains 51717 records with 17 features.

**Checking the columns in the dataset.**

**Columns description**

1. **URL** contains the url of the restaurant on the Zomato website 2. **address** contains the address of the restaurant in Bengaluru

**3. name** contains the name of the restaurant

**4. online\_order** whether online ordering is available in the restaurant or not **5. book\_table** table book option available or not

**6. rate** contains the overall rating of the restaurant out of 5

**7. votes** contain the total number of rating for the restaurant as of the above-mentioned date

**8. phone** contains the phone number of the restaurant

**9. location** contains the neighbourhood in which the restaurant is located **10. rest\_type** restaurant type

**11. dish\_liked** dishes people liked in the restaurant

**12. cuisines** food styles, separated by comma

**13. approx\_cost(for two people)** contains the approximate cost for a meal for two people

**14. reviews\_list** list of tuples containing reviews for the restaurant, each tuple **15. menu\_item** contains a list of menus available in the restaurant **16. listed\_in(type)** type of meal

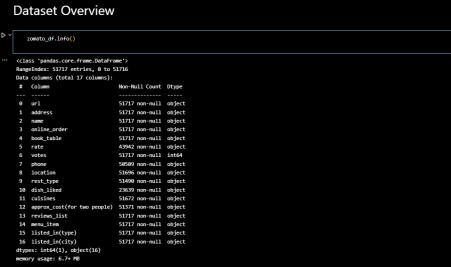
**17. listed\_in(city)** contains the neighbourhood in which the restaurant is listed.

**Understanding Overview of features**

● How the information is stored in a DataFrame or Python object affects what we can do with it and the outputs of calculations as well. There are two main types of data those are numeric and text data types. ● Numeric data types include integers and floats.

● Text data type is known as Strings in Python, or Objects in Pandas. Strings can contain numbers and / or characters.

● For example, a string might be a word, a sentence, or several sentences. ● Will see how our dataset is, by using **info ()** method.



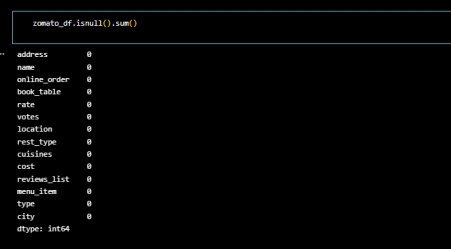
● As you can see in our dataset, except ‘votes’, all other features are categorical data, but it is not necessary that all the continuous data which we are seeing has to be continuous in nature. There may be a case that some categorical data is in the form of numbers but when we perform info () operation we will get numerical output. So, we need to take care of those type of data also.

**Checking for null values in the dataset**

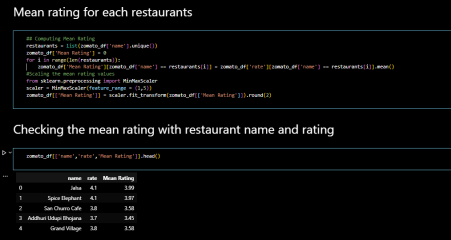
****

**Data cleaning as our dataset contains null values and some special characters**

****

**Checking for null values after cleaning & data Processing **

**Checking mean rating with restaurant name and rating for each restaurant using below line codes**

****

**We will be using the ‘Review’ and ‘Cuisines’ feature in order to create a recommender system. So we need to prepare and clean the text in those columns.**

**Operations performed: Lower Casing, Removal of Punctuations, Removal of Stop words, Removal of URLs, Spelling correction**

****

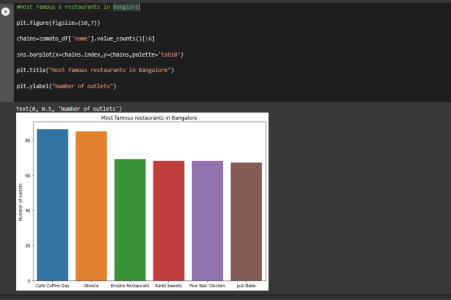
**Milestone 3: Data Visualization**

Data visualization is where a given data set is presented in a graphical format. It helps the detection of patterns, trends and correlations that might go undetected in text-based data. Understanding your data and the relationship present within it is just as important as any algorithm used to train your machine learning model. In fact, even the most sophisticated machine learning models will perform poorly on data that wasn’t visualized and understood properly.

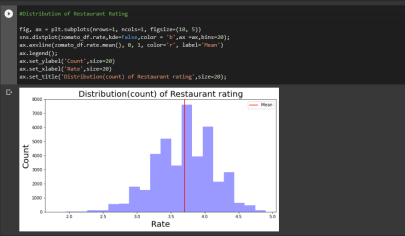
To visualize the dataset, we need libraries called Matplotlib, Seaborn. The Matplotlib library is a Python 2D plotting library which allows you to generate plots, scatter plots, histograms, bar charts etc.

Let’s visualize our data using Matplotlib and seaborn library.

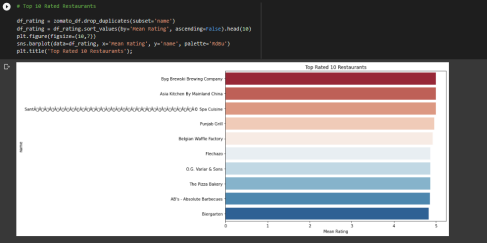
At first, we will be plotting a bar plot using matplotlib for showing the top 6 restaurants in Bangalore by value counts.



Checking the distribution of restaurant rating, for that we are using distplot from seaborn library.

And we can infer that most of the restaurants in Bangalore have rating above 3.5.

Visualizing top 10 rated restaurants in Bangalore. For that we are again using barplot from Matplotlib library.



Visualizing two word frequencies for cuisines, using barplot from seaborn library. 

Here we can see the Top favourite cuisine among people of Bangalore is ‘North Indian’, ‘Indian Chinese’ and ‘Fast food’.

**Milestone 4: CONTENT-BASE RECOMMENDER SYSTEM**

**Activity:1 TF-IDF Matrix (Term Frequency — Inverse Document Frequency Matrix)**

TF-IDF is a statistical method of assessing the meaning of a word in a given document. Now we use TF-IDF vectorization on the dataset.

Cosine similarity is a metric used to determine how similar the documents are irrespective of their size. Calculating the cosine similarity of each item with every other item in the dataset.



Here, the tf-idf matrix is the matrix containing each word and its TF-IDF score with regard to each document, or item in this case. Also, stop words are simply words that add no significant value to our system, like ‘an’, ‘is’, ‘the’, and hence are ignored by the system.

**Calculating Cosine Similarity**

****

And in the last line of code, we are calculating the cosine similarity of each item with every other item in the dataset. So we just pass the matrix as an argument.

**Activity :2 Creating Recommendation system**

****

**Querying recommendation for 4 Restaurants:**

**For Restaurant ‘Red Chilliez’**

****

**For Restaurant ‘Cinnamon’:**

****

**For Restaurant ‘Spice up’:**

****

**For Restaurant ‘Desi Doze’:**

****

**Milestone 5: Application Building**

**Activity 1: Create an HTML File**

We use HTML to create the front end part of the web page.

Here, we created 2 html pages- index.html, web.html.

index.html displays home page.

web.html accepts the values from the input and displays the prediction. For more information regarding HTML refer the link below

**https://www.w3schools.com/bootstrap/bootstrap\_forms\_inputs.asp**

● We also use JavaScript-main.js and CSS-main.css to enhance our functionality and view of HTML pages.

**○ Link :https://www.w3schools.com/css/**

○ **https://www.w3schools.com/js/DEFAULT.asp**

**Activity 2: Build python code**

**●** Let us build flask file ‘app1.py’ which is a web framework written in python for server-side scripting. Let’s see step by step procedure for building the backend application.

**●** App starts running when “\_\_name\_\_” constructor is called in main. **●** render template is used to return html file.

**●** “GET” method is used to take input from the user.

**●** “POST” method is used to display the output to the user.

**Importing libraries**

****Libraries required for the app to run are to be imported.

**Creating our flask app and loading the newly created dataset**

Now after all the libraries are import we will be creating our flask app with the updated dataset



**Routing to the html Page:**

Basically, we give routes of our html pages in order to show case the UI. By giving the routes the built code in the html page is connected to our flask app. This is how a UI can be built and showcased.

We are routing the app to the html templates which we want to render.

Firstly, we are rendering the home.html template and from there we are navigating to our prediction page that is indexnew.html



Lastly, we run our app on the local host.



Here we are running it on localhost:5000

**Activity:3 Run The app in local browser**

****

● Open anaconda prompt from the start menu

● Navigate to the folder where your python script is.

● Now type “python app.py” command

● Navigate to the localhost where you can view your web page **Showcasing The UI**

This is the home main page that describes the project and summarizes it.

I. Checking recommendation for the restaurant: ‘Jalsa’

This is the prediction page where we will provide a restaurant name for which we will get the top recommended restaurants, which based on cuisines, mean rating (out of 5), cost in thousands.

II. Checking recommendation for the restaurant ‘Cinnamon’ And here is the recommendation Finally, the prediction for the given restaurant inputs is shown.