Restaurants Recommendation System

Authors: Rishabh Sureshkumar Shah

Department of Information Systems, California State University Los Angeles

CIS5100 IS/IT Architectures

rshah39@calstatela.edu, hthakka4@calstatela.edu

***Abstract*- The purpose of this project is to develop various recommender systems, analyze them, and get useful information from them. We used information from eateries in Bangalore, India that are included on the well-known food delivery and review website Zomato. Using the knowledge we gained during the exploratory stage of our project, we developed a method for developing recommender systems that makes use of the advantages that both datasets have to offer. The reviews of customers in India (about particular food products). The presentation of a recommender system technique tries to give users restaurant recommendations based on their preferences. In addition to a discussion of performance, model trade-offs, and workaround options, several alternatives are put forth.**

Keywords: Recommendation system, Zomato, Analysis, Machine Learning, TF-IDF, Collaborative Based-Filtering, Content Based-Filtering.

1. Introduction

A recommender system is a kind of information filtering system that attempts to anticipate the "rating" or "preferred" a user would assign to an object. Recommender systems help users make informed decisions by collecting information about their preferences to suggest similar goods, services, items, etc... Results are usually based on historical data or users with similar preferences. There are many applications for recommender systems, some of which include content recommenders for social media platforms, playlist generators for video and music services, and online store product recommendations. Systems that recommend meals and locations to eat based on a user's preferences, review behavior, ratings, and other factors are also very common. Recommendation algorithms are also used by businesses like YouTube, Instagram, Meta, Netflix, Google, and others. Recommendation systems are widely utilized and have been employed in research to filter pertinent information for a user from vast collections of data in an exponentially expanding world of big data. Due to significant technological breakthroughs that have made it possible for us to become hyperconnected during the past ten or so years, these extremely complex systems have seen significant growth. These days' smartphones feature high bandwidth information displays and very quick internet, therefore large organizations have resorted to information filtering systems to keep consumers interested, offer a more individualized experience, and boost platform satisfaction overall.

A decent recommendation system should consider the context of the diners in addition to the key restaurant characteristics. Despite the abundance of context-aware restaurant recommendations, the most of them just consider the surrounding area.

A user's final recommendations may be influenced by a variety of features that can make recommender systems quite complex. Some of these factors could range from having a lot of context to having very little context. Others are more sophisticated and may consider the previously mentioned factors in addition to the app's specified filters, button pushes, and history search words. Some recommendation systems are based on a limited collection of features, such as historical location data.

NumPy is a basic scientific computing package for Python. NumPy stands for numeric python, is a Python library for computing and processing multidimensional and linear array elements. Travis Oliphant created the NumPy package in 2005 by incorporating the functionality of the progenitor module Numeric into another module called Numarray.

NumPy is an important library for Data Science, Machine learning, Image processing, and Scientific and engineering computing. Numpy makes many mathematical operations: Vector-Vector multiplication, Matrix-Matrix, and Matrix-Vector multiplication, Element-wise operations on vectors and matrices (i.e., adding, subtracting, multiplying, and dividing by a number), Reduction, statistics

Benefits to use NumPy:

More speed: NumPy uses algorithms written in C that complete in nanoseconds rather than seconds.

Fewer loops: NumPy aids in the reduction of loops and prevents you from becoming confused by iteration indices.

Clearer code: Without loops, your code will look more like the equations you're trying to solve.

Better quality: There are thousands of contributors working to keep NumPy fast, friendly, and bug-free.

Pandas is the most popular data manipulation and data analysis software library for the Python programming language. It is an open-source library that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures. This library is built on top of the NumPy library.

Pandas were developed by Wes McKinney in 2008. Chang, was the second major contributor to the library in 2012. The panda's most recent version, 1.5.0, was made public on September 19, 2022.

As data scientists and programmers who are familiar with the R programming language for statistical computation are aware, DataFrames are a means of storing data in grids that are easily viewed. This means that Pandas is chiefly used for machine learning in the form of DataFrames.

Pandas generally provide two data structures for manipulating data,

Series: Pandas Series is a one-dimensional labeled array capable of holding data of any type (integer, string, float, python objects, etc.). The axis labels are collectively called indexes. Pandas Series is nothing but a column in an excel sheet. Labels need not be unique but must be a hashable type. The object supports both integer and label-based indexing and provides a host of methods for performing operations involving the index.

DataFrame: Pandas DataFrame is a two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the data, rows, and columns.

Pandas is well suited for working with several kinds of data, including:

Tabular data with heterogeneously typed columns, as in an SQL table or Excel spreadsheet

Ordered and unordered time series data

Arbitrary matrix data with row and column labels

Advantages to data scientists and developers:

Easy handling of missing data in both floating-point and non-floating-point data

Fast and efficient for manipulating and analyzing data.

Data from different file objects can be loaded.

Data set merging and joining.

Flexible reshaping and pivoting of data sets

Provides time-series functionality.

Hierarchical labeling of axes

2. Related Works

We have explored some information from previous research as reference material about the pros and cons that already exist.

A restaurant is a public place that provides one dish or various dishes. Currently, many applications contain information about restaurants, for example, name of restaurants, menus, food prices, ratings, reviews, etc. The system that is popularly used in restaurants is the recommender system [1].

Various studies related to the recommender system have been carried out. A recommender system is an effective way to help users to get information that is useful and to their user interests [2]. With the increasing amount of information on the internet, recommender systems can solve the problems caused by increasing information rapidly [3][4]. For example, YouTube, Google, Netflix, Meta, and Instagram use recommendation systems in their respective areas.

3. Analysis of The Project

Before reaching the conclusion of the project the authors did some exploratory data analysis on the dataset to get detailed insights from the data. Here are some of the results of the analysis that the authors did on their respective dataset.

The first analysis that the writers did was to locate the location of the restaurants that were available in the dataset. This figure helps to locate the restaurants and get an idea in which area there are more and fewer restaurants. This analysis was done with the help of the "folium" library of python which allows us to see the results in the form of a map. The map that we can see is of Bangalore city and the red spots are because of the use of the heatmap. The primary purpose of Heat Maps is to better visualize the volume of locations/events within a dataset and assist in directing viewers toward areas of data visualizations that matter most.

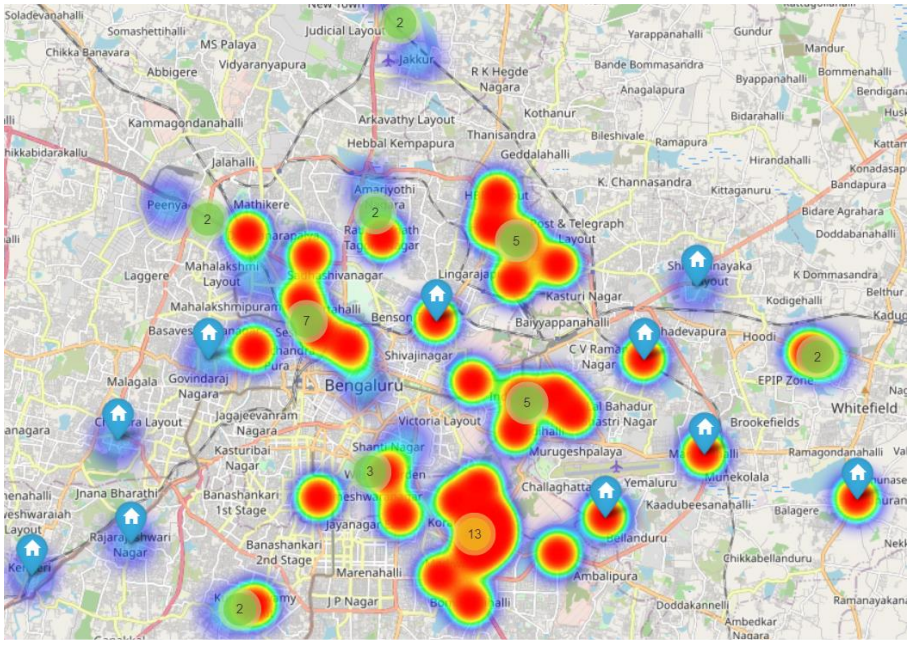

Figure - 1

Next analysis was done to visualize which restaurants give the option to deliver food at home.

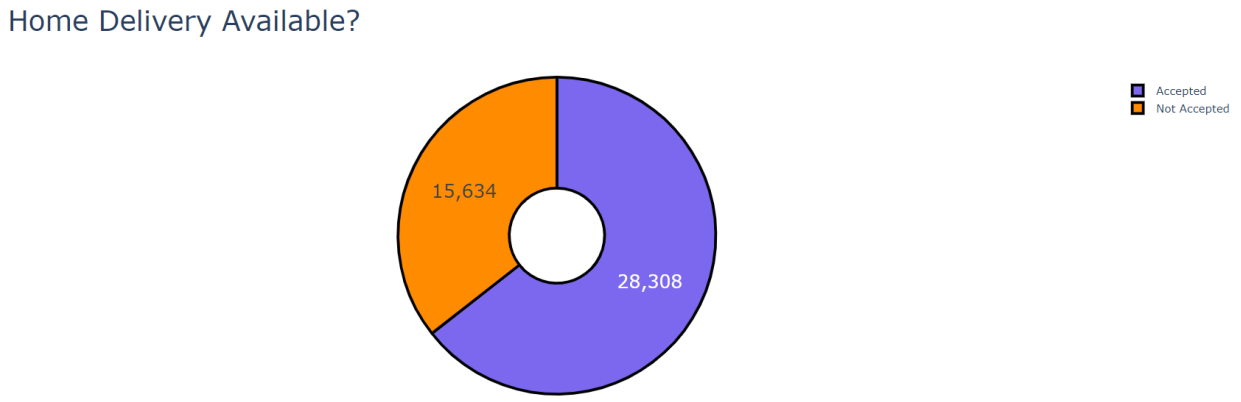

Figure - 2

After that, the researchers analyzed to visualize which restaurants provide an option to book a table. For that, the results are as below in the figure.

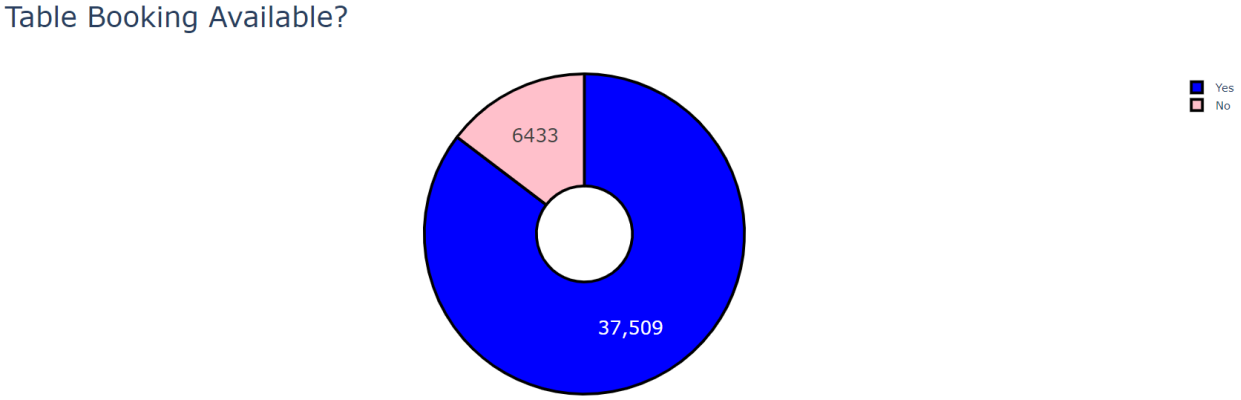

Figure - 3

After visualizing the booking option available or not, the author moved on to check the most and least popular cuisine options available in Bangalore. Here are the results for that.

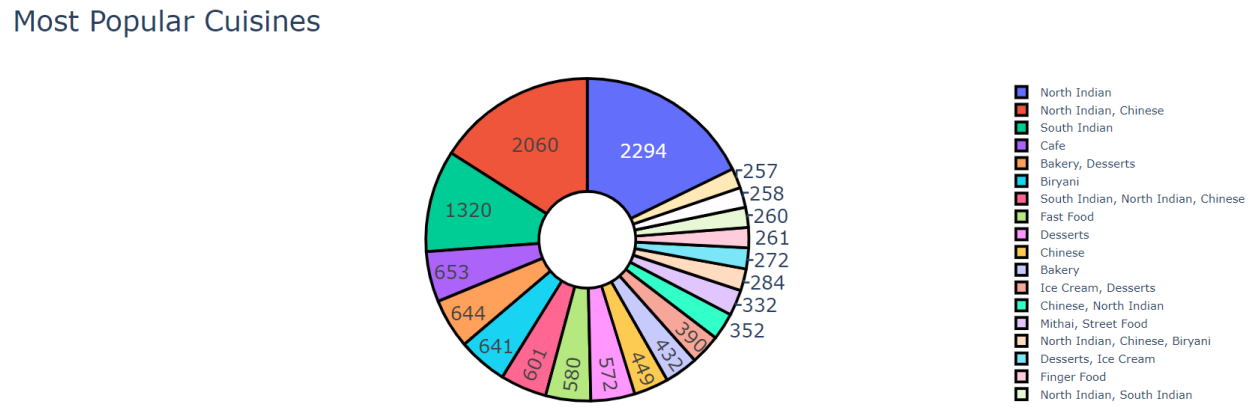

Figure - 4

4. Implementation Flowchart

4.1 Implementation of Restaurant Recommendation System

Figure - 5

4.2 Types of Information Filtering Systems

We plan to include a few different variations of collaborative and content-based filtering information systems for our recommender systems. Content-based filtering provides recommendations to users based on the historical context (restaurant type, cuisines, home delivery, table booking, etc.) that a user consumed. Collaborative-based filtering provides recommendations to users based on similar habits identified among groups of users; recommendations can then be made…

Query Based

Collaborative Based

Cite using APA

Put

4.3 Dataset Background

Coming to the dataset of our project, a location that we selected for recommendation is "Bangalore" city (a.k.a Silicon Valley of India). The Zomato dataset had about 52,000 rows and 17 features which included the features as IDs, ratings, reviews, rest_type, dishes_liked, cuisines, etc. The writer also dealt with null values in the dataset.

4.4 Flow Chart

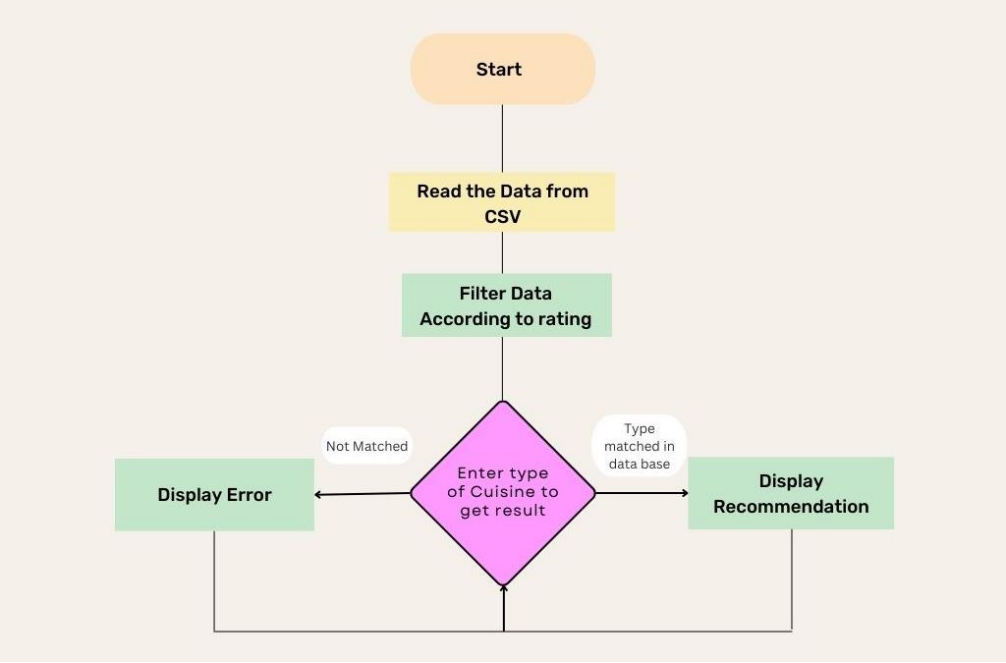

Figure - 6

5. Result

We have successfully implemented the recommendation system required according to the Query which can be shown in the figures below. As we enter the type of cuisine present in this data set it filters the data with that particular cuisine and gives us the output with the best ratings

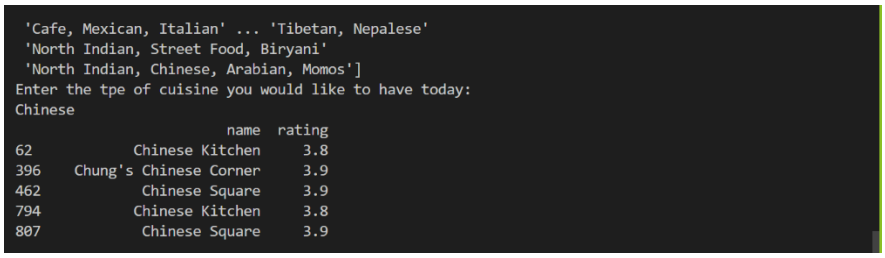

Figure - 7

Whenever there is any query that is not present in the dataset it gives us an output with an empty data frame and asks us to input another query to get the desired output.

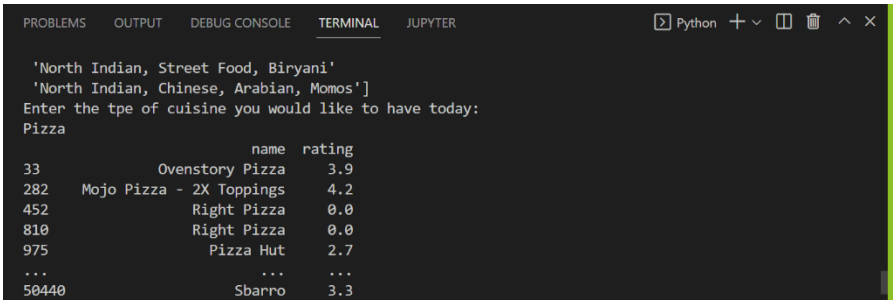

Figure – 8

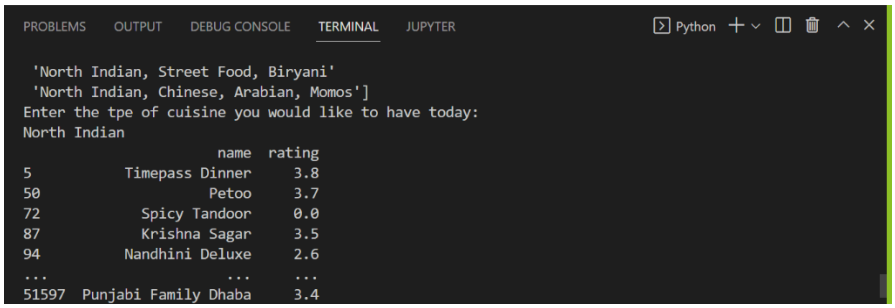

Figure - 9

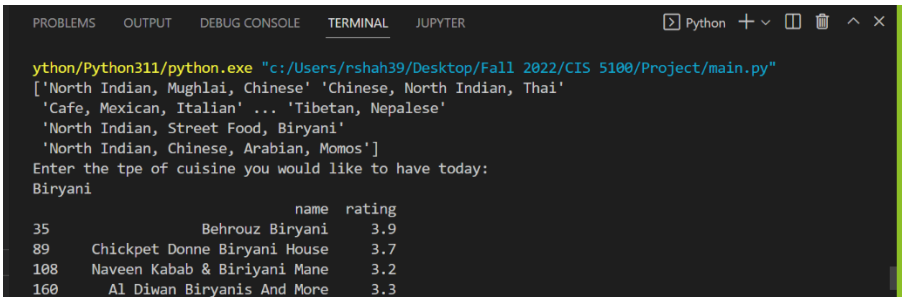

Figure - 10

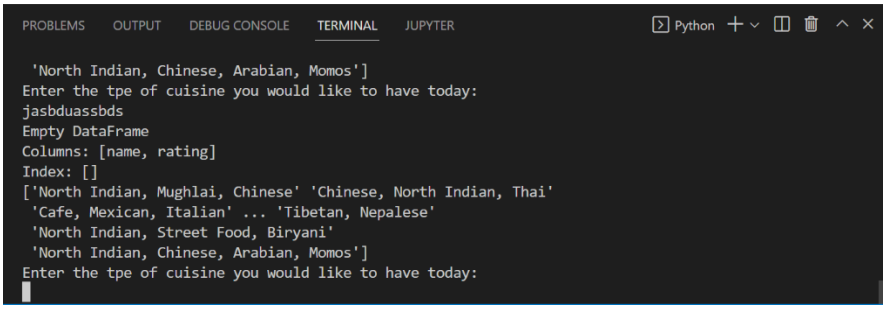

Figure - 11

6. Future Scope

In this project, we have successfully analyzed and processed data to give the best possible restaurant based on the type of cuisine we want to have. For an instance, if we want to have pizza, so we get a recommendation of the best restaurant known for pizza for a particular data set. Furthermore, we are also planning to analyze users' locations, history of the order, to suggest to them the best restaurant in that area according to personal requirements. We have already got the longitude and latitude values for this data set, to add these special features to this system.

7. Conclusions

We can conclude from our project that we can utilize the data present to make predictions according to our needs which is the query here in our case. We can further do this automatically as the user with the app might be searching for a place to eat. The performance of this model is dependent on the size of the dataset and the number of factors we consider to predict the output. As we take more amount of information into consideration the need for computational power increases and the complexity of the code also increases as there can be a number of loops looking for more and more information to get the possible output.

8. References

[1] Gupta A and Singh K 2013 Location-based personalized restaurant recommendation system for mobile environments Advances in Computing, Communications and Informatics pp. 507-511 IEEE The 2nd International Conference on Data and Information Science IOP Conf. Series: Journal of Physics: Conf. Series 1192 (2019) 012023 IOP Publishing doi:10.1088/1742-6596/1192/1/012023 7

[2] Zeng J, Li F, Wen J, and Hiroka S 2016 A restaurant recommender system based on user preference and location in mobile environment Advanced Applied Informatics pp. 55-60 IEEE.

[3] Jonnalagedda N, Gauch S, Labelle K, and Alfarhood S 2016 Incorporating popularity in a personalized news recommender system PeerJ Computer Science 2 e63

[4] Baizal Z K A, Widyantoro D H, and Maulidevi N U 2016 Factors Influencing Users Adoption of Conversational Recommender System Based on Product Functional Requirements 1575-1585 TELKOMNIKA

[5] Arigi L R H, Baizal Z K A, and Herdiani 2018 A Context-aware recommender system based on ontology for recommending tourist destinations at Bandung Journal of Physics: Conference Series 971 IOP Publishing.

[6] Baizal Z K A, Rahmawati A A, Lhaksmana K M, Mubarok M Z, and Qadrian M 2018 Generating Travel Itinerary Using Ant Collony Optimization 16(3) Telkomnika.

[7] Wang B, Huang J, Ou L, and Wang R 2015 A collaborative filtering algorithm fusing user-based, item-based and social networks pp. 2337-2343 IEEE.