**Project title : Restaurant Recommendation System**

**Team Name: Trojan Horse**

|  |  |
| --- | --- |
| Team member Name | SRN |
| Jahnavi A B | PES2UG19CS157 |
| Nikhil V R | PES2UG19CS258 |
| Raghav S K | PES2UG19CS312 |
| Vivek S D | PES2UG19CS461 |

Github Link for the jupyter notebook:

<https://github.com/nikhilvr13/Data-Analysis-Project---Trojan-Horse>

1. Dataset Name and Description.

Zomato Bangalore Restaurants

The dataset contains 17 variables and 51,717 rows all of which were scraped from the Zomato website. The dataset contains details of more than 50,000 restaurants in Bengaluru in each of its neighborhood.

The dataset also contains reviews for each of the restaurant which will help in finding overall rating for the place.

For each of the neighborhood and for each category their the dataset contains onlineorder, booktable, rate, votes, phone, location, resttype, dishliked, cuisines, approxcost(for two people), reviewslist, menu\_item.

1. Problem statement

Build a restaurant recommendation system based on the various inputs from the user like location, type of dish, cost etc.

1. EDA and Visualization
   * How many rows and attributes?

51,717 rows and 17 attributes

* + How many missing data and outliers?

A picture containing text

Description automatically generated

* + Any inconsistent, incomplete, duplicate or incorrect data?

The attribute ‘menu\_item’ had to be dropped as more than half of the values had incomplete data(empty list).

* + Are the variables correlated to each other?

Yes the variables are correlated to each other like cost and rating, dish\_liked and rating etc.

* + Are any of the preprocessing techniques needed: dimensionality reduction, range transformation, standardization, etc.?

Dimensionality reduction by dropping some columns .

Data Cleaning techniques like imputing the missing data, handling the null values were needed.

* + Does PCA help visualize the data? Do we get any insights from histograms/ bar charts/ line plots, etc.?

Plotting of boxplots, histograms, pie charts, heatmaps helped to get insights about the relation between price range of the restaurants and rating, the types of cuisines and the types of restaurant that people choose the most etc.

Chart, box and whisker chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, pie chart

Description automatically generated

1. Link for google sheet

<https://docs.google.com/spreadsheets/d/1cAeaM_zUKacWOOdwy1EvwYfxM52Rq8x5iGeFBNuGEpI/edit?usp=sharing>

1. Literature Survey ( Summarize)

- **a type of recommender system that attempts to guess what a user may like based on that user's activity**. Content-based filtering makes recommendations by using keywords and attributes assigned to objects in a database (e.g., items in an online marketplace) and matching them to a user profile.Implementing a content based recommender system makes the recommendations highly transparent to the user

- Collaborative filtering addresses the limitations of content based filtering and uses similarities between users and items simultaneously to provide recommendations. This allows for serendipitous recommendations

- These filtering methods can be used to build a recommendation system based on the user’s choices

- The k-means clustering is fastest and is a very good clustering algorithm.

-In user based collaborative filtering , excluding the user attributes in calculations gives better performance because preference information from user attributes is less so the resulting calculation becomes less accurate.

**Recommendation system for mobile environments**

This study proposes a model that combines localization, personalization and content-based recommendation in a dynamic and ubiquitous environment.

**-Recommendation System for User Preference and Services Based on Rating and Amenities**

The research work proposes the strategy of Natural language Processing (NLP) machine learning algorithm by examining the behavior of users, employing the text data and ratings given by the users.

**-Recommendation System Based on Item and User Similarity**

This research produced a recommendation system based on item and user similarity. The approach is built by combining features derived from content based filtering and collaborative filtering approaches.

1. Your Plan

Building a recommendation model based on content based learning, collaborative filtering , simple k means algorithm for clustering. Training the model on a part of the dataset and testing the same.

1. References

<http://cs229.stanford.edu/proj2014/Ashish%20Gandhe,Restaurant%20Recommendation%20System.pdf>

<https://ijesc.org/upload/40157934dada8de435e6326872494608.Restaurant%20Recommendation%20System%20Using%20Customers%20Data%20Analysis.pdf>