Final Project Report

Sustainability of restaurants based on location demographics.

The University of Michigan – Dearborn

Teja, Pruthviraj, Vinayaka, Deepak
CIS 5570 – Intro to Big Data
Prof. Mohamed Abouelenien, Winter 2020

Abstract

The core idea of our project is to analyze the aspects that affect the growth of restaurants at different locations and to address questions regarding restaurant feasibility and sustainability. By incorporating user preference functionality from the range of food cuisines in restaurants, our recommender system adopts a consumer preference paradigm and uses local restaurant knowledge to suggest the results. The outcome of this experiment is that the current framework for recommending a restaurant will easily take advantage of user interest and local restaurant knowledge to suggest the person suitable restaurants based on his preferences.

Introduction

Metropolitan food culture is vibrant and intriguing. Restaurants include cuisines from all over the world. You name it, and urban towns have it, "Delivery, Dine-out, Pubs, Bars, Drinks, Buffet, and Dessert bars". The restaurant business is evolving and improving day in and out. The food industry will never be saturated as there is a continuous ongoing demand every single day. And every day, new restaurants appear. However, it has become difficult for them to compete with already established restaurants. With widespread digitalization, this industry is generating huge amounts of data every day. Solutions to this utilizing the generated data will help new restaurants in deciding their theme, menus, cuisine, the cost for a particular location. The dataset also contains reviews for each of the restaurants which will help in finding the overall rating for the place. For example, we can talk about the Metropolitan city of Bengaluru, India. Here, restaurant data has been collected from a popular food delivery App viz "Zomato." This Zomato dataset is used for analyzing the demography of the location and other factors affecting a restaurant's success. Most importantly, it lets new restaurants determine their theme, their menus, their cuisine, and their costs.

Related work

[1] Suggests a restaurant recommendation engine and assumes people often tend to go to the nearer restaurants. The system utilizes the characteristics of the visited restaurants to follow a customer favorite layout. This pattern of choice can be used easily to make recommendations to different users. The suggested system [1] often utilizes customer and restaurant position details to generate the recommendation output.

[2] Explores the efficiency of a content-based program through the influence of feature selection. This paper explores specific solutions to advice such as Collaborative recommendation: Memory-based and Model-based filtering, and Content-Based evaluation. [2] establishes a hybrid recommendation framework for considering the user criteria and restaurant features in order to construct a recommended framework.

Methodology

The project is performed in an agile model and split into stages. The first process consists of project planning and resource acquisition. The second stage involved concept assessments and the study of viability. Thirdly, the results are compiled and the methodologies to be used are determined. The fourth step is data cleaning and exploratory data analysis. The fifth and final step is building a recommendation system.

- Data Cleaning: The data file we have referred to in our project has 17 columns , the rows had data with some anomalies , like having empty cells for certain columns , Null values against few columns. To use the csv data for further process , the anomalies needed to be removed by cleaning the data.
- Exploratory Analysis is performed on the data where we analyze data by plotting it against various parameters like number of restaurants, locality, price and many more factors.
- Content Based filtering is based on content , it uses item features to recommend certain things close to what the user likes, based on their choices or explicit feedback.
  Steps involved in here are feature extraction, computing the tf, idf and cosine similarities.
- Regression Analysis is used to discover the statistical relationship between different variables combined against the ratings. Regression techniques used are Simple Linear regression , Random forest Regressor ,and Decision Tree Regressor.
- Clustering of data is done by extracting the feature matrix , using k centroids and adjusting them so that they become the centre of gravity then measure the distance with respect to every point by reclustering. This procedure is repeated till data points stop changing the clusters.

Experimental Discussion

Content-Based Recommendation
Suggest similar items on the grounds of a specific item. In such suggestions, this program uses metadata objects, such as locality, cuisine, rating, etc. The general concept behind these systems is that if a person likes a particular item, he or she will also like an item that is similar to it.
For example, in our data, we choose the Vijayanagar location for a particular restaurant and based on that restaurant's cuisine e.g North Indian and South Indian and Chinese and recommend restaurant with similar cuisines in that locality.

K-means Clustering

We applied the clustering method to analyze the clusters with similar attributes with respect to the variables. It is important to first prepare the dataset to perform clustering on it. So, we perform a feature matrix, choosing the numerical data types followed by standardizing the features before clustering. We have used the StandardScaler method and fit transformation to get a good optimal value of K=4 value (Elbow & Silhouette Method) which explains us more about the behavior of the variables.

Contribution

While a team, each participant has contributed to each step of the project. Every phase the meetings are analyzed and reviewed by all team members. For the tasks, we split into sub-teams for fulfilling the different tasks in the project. Teja is involved in the Data Acquisition, Regression Analysis and literature survey. Pruthviraj is involved in data cleaning, literature survey and clustering. Vinayaka is involved in Exploratory Data Analysis, Content based recommendation and literature survey. Deepak is involved in Data Cleaning, Exploratory Data Analysis, Simple recommendation and literature survey.

Conclusion

Recommendation is an important way of providing users with customized and useful knowledge. This recommendation system used a user preference model to recommend similar Cuisines in a locality. As future work, we can apply collaborative filtering by gathering user-restaurant data which drives better recommendations based on user preferences.

References

1. Jun Zeng, Feng Li, Haiyang Liu, Junhao Wen, Sachio Hirokawa"A Restaurant Recommender System Based on User Preference and Location in Mobile Environment" in 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)
2. Yassine Afoudi, Mohamed Lazaar, Mohamed Al Achhab "Impact of Feature selection on content-based recommendation system" in the proceedings of IEEE 2019.
3. Jure Leskovec, Anand Rajaraman, Jeff Ullman, Accessed April 5 2020 "Mining Massive Datasets" http://mmds.org/