Food Order and Analysis

Ishan Kaushal1 ,Sourabh Shubham2 ,Miss Megha Sharma3

1, 2 CSE Student, 3 Asst. Professor/CSE SRM Institute of Science and Technology Delhi, NCR

1.Abstract

"Food Order and Analysis" is an interesting and one stopplatform to know about the frequently ordered and mostacclaimed food around us.This will be done using interactiverepresentation. This project is all about using Data Analysis methods to analyse the data. This project conveys the useful findings about most famous cuisines and food interests of an individual as well as group of people by analyzing the history of the food ordered. Many different food websites, food bloggers, hotels and restaurants can take advantage of this project.This will help them in expanding their business on the basis of the outcomes of this project. Intention behind this project is spreading more and more of hidden secrets and conclusions of the most common food related happenings around us from decades. This project will help anyone to open and grow its business related to restaurants in any part of the country as it analysis the data state wise,it can also be useful it we want to create new fusion dishes.

2 Introduction

Data mining is that part of data science in which we use our data to collect some useful information from it .This can also be termed as turning the undistributed data into some kind of useful information .

Our project is also using the concept of Data Visualization in which we are drawing out different types of graphs, histograms, etc to draw out beautiful information in a very interactive and interesting form .

Data visualization allows someone to easily understand the result from a data set which is very complex to understand

It creates a very clear understanding between the user and the consumer

2 Literature Survey

Hessah review system(sentiment analysis) is used to differentiate between good and worst restaurant by checking people reviews and dividing into good or bad review ,mobile based app ,users phone location for accuracy to find nearby restaurant,they also uses image to show food quality, Arabic language support available.Sentiment analysis can be expanded to different languages.

PkaypreakUnganything related to restaurant, booking, review, customer support,chef info,work related to take out and delivery options.Preference and Services Based On Rating and Amenities R.M. Gomathi uses tripadviser data for its improvement,NLP - Natural Language Processing is used to differentiate between good and bad reviews,user is asked to select location and features and based on that restaurants are shown.

Recommendation based on Customer Preference.

DitditNugeraha uses fuzzy-logic and bubble-sort for restaurant recommendation,three selected parameters; interest, location, and rating .Adding more parameters (budget),expanding to different countries as for now only available in Indonesia

VindyaLiyanage tries to address the deficiencies such as long waiting time ,menu discrepancies etc and provide accurate services to customer.The concept is implemented as a mobile application using latest IT concepts such as Business Intelligence, Data Mining, Predictive Analysis and Artificial Intelligence.Unique food recommendation and it's order for each customer is generated by analyzing their social media information

3 Research Gap

The above-mentioned papers and after going through some others papers as well, literature review showed that there are numerous research papers having similar independent feature set. Some of the papers only used complex situations for determining the various conclusions and all the conclusions drawn were very general .So it is important for us that we create a project which is person or business specific. The food quality should also play an impact role in determining our conclusions and then applying different algorithms will be helpful

earch Question

ing are the question will be answered by the end:

Type of food combinations work ?

type of people like which kind of food ?

oes the pricing of food impacts the no of times that ation being used ?

ific food consumed in any area specific for that certain

oes the business owner setup his/her business based on d combinations?

ective of Study

ing are the objectives of study:

1. To create a system that can identify and suggest restaurants food items they can create and sell

2. To explore and compare the best dishes that are being sold in that particular region
3. To give restaurants the edge over other and make their own speciality by create interesting menu
4. To help someone establish a business on the basis of food combinations, pricing for two and many features

7 Collection of Sample

Required dataset was searched on different websites which provide authentic data likegoogle dataset, kaggle, After collecting the sample data set we cleaned the data set and then basically applied simple data visualization techniques to get the result.Upon finalizing the source, we looked for discrepancies in the dataset. They were further resolved to make the dataset suitable for model training. At last different models were used and comparison was made between them to arrive at final mode

8. Methodology and Definition of Variables

In our first phase of project, we collected information from different websites which are famous for datasets, namely google dataset, kaggle, uci machine learning repo etc. Upon finalizing the source, we looked for discrepancies in the dataset. They were further resolved to make the dataset suitable for model training. At last different models were used and comparison was made between them to arrive at final model

1.Data Collection

2.Data visualization

3.Data selection

4.Data preprocessing

5.Data transformation

6.Modeling

7.Model evaluation and testing

The different data features were having different scales. For example,food order data was provided to us for around 21 countries and out of that we only choose 1 for simplicity of our project. To bring them to a comparable scale standard scaler was used.During this stage, we tried to create as many as possible graph to check and verify the most accurate predictions and to get better performance result.We used three different

8.Model Deployment

For this project dataset collection was not easy as only one dataset waspresent over the internet which had many missing values. Thus, we were left with only option to collect and organize the data from different countries. So, we used the Zomato dataset present on Kaggle for this purpose. We used 'The more, the better' approach and collected data over a range of five years. We visualized the data to look for noise, missing values, dependency, range of values of a particular feature. Visualization was performed for only certain features.After we had collected all the information, we had to filter them. After going through various research papers, we filtered out 8 features from 21 features to begin with our project. Correlation matrix and heatmap was used to visualize and carefully pick the most responsible features. Preprocessing was done to convert the raw data so that it fits into the machine. Structured and clean data was made to get more precise results by applying data visualization techniques.

graphs namely barplot , scatterplot, and heat map. After comparison scatterplot was selected as it produces the most accurate mean and performance. Cross validation was also used in some cases were results showed overfitting. This stage covers putting model into use. Once our model was finalized, we use jupyter notebook to extract the model and then use it with ease.

Data Columns

```
data.columns
```

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes'],
      dtype='object')
```

HeatMap

```
In [19]: plt.figure(figsize=(10,8))
         sd=sns.heatmap(c.corr(),annot=True)
         plt.title("HeatMap for India's Dataset")

Out[19]: Text(0.5, 1.0, "HeatMap for India's Dataset")
```
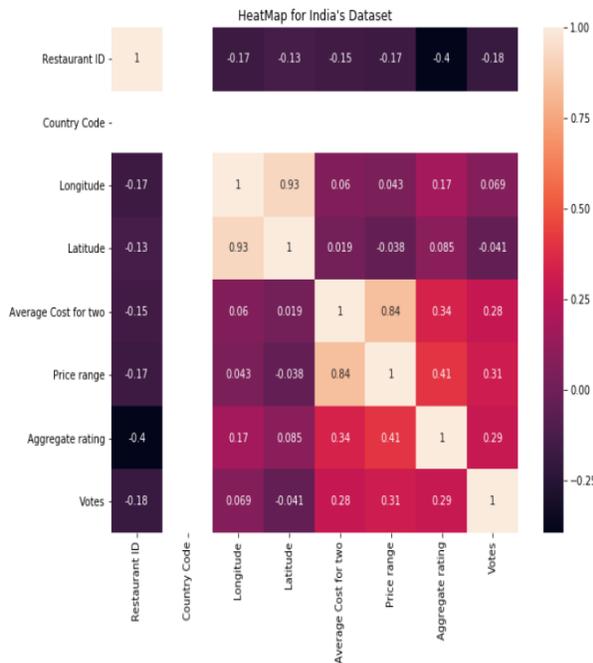

Average cost

```
plt.figure(figsize=(7,5))
ax = sns.scatterplot(x="Average Cost for two", y="Cuisines",data=c.head(30))
plt.title("Scatter plot to show the average cost for two per cuisine")

Text(0.5, 1.0, 'Scatter plot to show the average cost for two per cuisine')
```
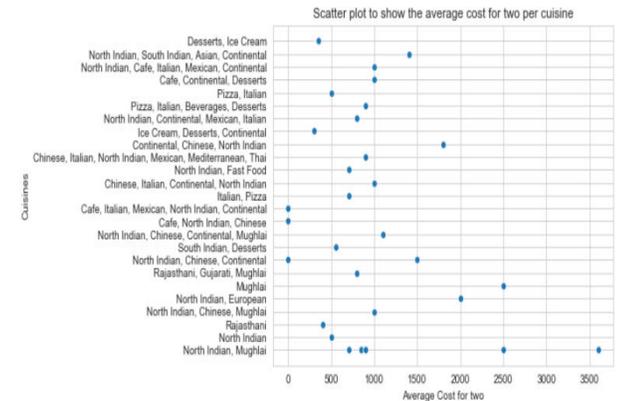

9 Steps of Analysis:

Following steps were followed in sequential manner in order to get the desired results which we were looking for:

1. After collecting the data, the first step was to analyze the data and filter out all the unnecessary features present in the dataset. The filtering was done after going through different research papers to know the meaning and importance of different features.

2. Different steps were followed to preprocess the data and make it suitable for model training.

3. Variables other than the desired country was removed.

4. Any kind of null values was removed and it was cleaned on the basis of negative or similar value

5. The first process in data cleaning was to remove the anomalies and missing values present in data. We used is null function to get the count of null values in a column.

6. Mean of the column was used to fill or replace the null values in each column where there were null values.

7. Datatype of each column was changed to numeric to allow different methods from NumPy and other modules to work on them.

8. After the previous steps, we sorted the data set in the required way .

9. Since the features had values of a very wide range it was required to bring them to a similar scale.

10. After completion of data cleaning and pre processing different regression algorithms were tried on the training data.

10 Findings of Study

- There are different types of research which had been done before we started our project ,but the complexities of such projects and papers was very high . So we thought of making a project which is business specific and food specific ,which would allow someone to strategize according to the specific need.

- Rather than using complex algorithms which were already used in earlier projects, we used simple data mining and data visualization techniques to get the desired results.

- We used different constraints(Budget, Location)and applied on the data set to bring out the result such as the most famous food for the area ,best combinations ,most consumed food, best average priced restaurants etc.

- According to different research papers there were many features which were highly responsible for Food analysis but the data corresponding to those features were mostly null or irrelevant.

11 Research Limitations

• Some features which seemed to be important after going through research papers had to be skipped because the dataset contained mostly null values for those features.

• Our project has a limitation that it can not be generalized and applied to all parts of country, rather it has specific usage for specific business .

• It uses simple mechanism so it may not perform brilliantly in complex situation. Many enhancements can be applied to this project and we could identify the specific dish which could not be found by our project .

• Finding the specific dish would have been ideal, but it is limitation for our project

12 Future Research Opportunities

• For now we can predict the cost for two person only if we know the values of input features but we will try to make a model which would give us some approximate value for every input feature.

• For countries having many sub provinces we will try to forecast the most sold food product of such individual states and provinces.

• Currently we only calculate the cost for 2 but later on we will try to calculate the best food item of a region in every state or provinces .

13 Practical Implications

Our project has a practical application in real life .For someone looking for setting up a business in a area where there is a lot of competition, specially in restaurants, then it is very important for the businesses to have something new and modern approach to succeed in such situations ,so our projects allows you to set up food combinations based on the study which is very scientific and logical . You should just see and

visualize and easily understand how our bar graphs and scatter plot gives different conclusions .The project will also allow someone to explore the hidden secrets and food related happening which occur and gets unnoticed

14 Conclusion

After completing all the steps, we are now in a position to answer the research questions. There are different types of research which had been done before we started our project ,but the complexities of such projects and papers was very high . So we thought of making a project which is business specific and food specific ,which would allow someone to strategize according to the specific need .Rather than using complex algorithms which were already used in earlier projects, we used simple data mining and data visualization techniques to get the desired results . It is evident from the graphs that different constrains must be used to find the best dish for two person. The constrains that were used are budget, location and review. In our analysis we got to know about twenty-one features, eight out of them were mainly responsible for our analysis. We can make the use of this study to predict the best food item for two of a region once we know the values of all the independent features. The last thing we wanted to do was to get the best option or combination of food for two people and we have successfully achieved our target, The model has been carefully trained to obtain a very high accuracy by removing all null values. Upon plotting and analyzing we found that the actual line almost coincided with the predicted plot. So, this estimation can be well used for the prediction

15 Reference

* H. Alsaaran, W. Alowain, D. Aldrees, S. Almadhi and W. Alqahtani, "Where to Eat? An Automatic Summarization System for Opinions in Restaurant Reviews," 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2020, pp. 1-5, doi: 10.1109/ICCAIS48893.2020.9096724.

* K. Chochiang, P. Ung and N. Bunsaman, "One Stop Restaurant Service Application," 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 2020, pp. 482-485, doi: 10.1109/ECTI-CON49241.2020.9158070.

* R. M. Gomathi, P. Ajitha, G. H. S. Krishna and I. H. Pranay, "Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862048.

* D. N. Utama, L. I. Lazuardi, H. A. Qadrya, B. M. Caroline, T. Renanda and A. P. Sari, "Worth eat: An intelligent application for restaurant recommendation based on customer preference (Case study: Five types of restaurant in TangerangSelatan region, Indonesia)," 2017 5th International Conference on Information and Communication Technology (ICoIC7), Melaka, Malaysia, 2017, pp. 1-4, doi: 10.1109/ICoICT.2017.8074654.

* Y. Shi, Q. Zhao, Y. Wang and J. Cao, "An Improved Restaurant Recommendation Algorithm Based on User's Multiple Features," 2018 IEEE 9th International Conference on Software Engineering and Service Science

(ICSESS), Beijing, China, 2018, pp. 191-194, doi: 10.1109/ICSESS.2018.8663850.

* H. Fidan, A. Teneva, S. Stankov and E. Dimitrova, "Consumers'Behavior of Restaurant Selection," 2018 International Conference on High Technology for Sustainable Development (HiTech), Sofia, Bulgaria, 2018, pp. 1-3, doi: 10.1109/HiTech.2018.8566405

* V. Liyanage, A. Ekanayake, H. Premasiri, P. Munasinghe and S. Thelijjagoda, "Foody - Smart Restaurant Management and Ordering System," 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Malambe, Sri Lanka, 2018, pp. 1-6, doi: 10.1109/R10-HTC.2018.8629835.

* Y. Jin, W. Zhang, M. Sun, X. Luo and X. Wang, "Neural Restaurant-aware Dish Recommendation," 2020 IEEE International Conference on Knowledge Graph (ICKG), Nanjing, China, 2020, pp. 599-606, doi: 10.1109/ICBK50248.2020.00090.

* K. Kaushal et al., "The E-Restaurant," 2016 Ninth International Conference on Contemporary Computing (IC3), Noida, India, 2016, pp. 1-5, doi: 10.1109/IC3.2016.7880205.

* A. A. Mustafa and I. Budi, "Recommendation System Based on Item and User Similarity on Restaurants Directory Online," 2018 6th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 2018, pp. 70-74, doi: 10.1109/ICoICT.2018.8528775.