National College of Ireland

Prediction Rating of Best Cuisines in Country Capitals Using Machine Learning Algorithms

MSc Research Project
Data Analytics

Priyanka Vesireddy
Student ID: X17152631

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Priyanka Vesireddy

Student ID: X17152631

Programme: Data Analytics **Year:** 2020

Module: MSc. Research Project

Supervisor: Jorge Basilio

Submission Due Date: 26 -05-2020

Project Title: Prediction Ratings of Best Cuisines in Country Capitals Using Machine Learning Algorithms

Word Count: 6598 **Page Count** 21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Priyanka Vesireddy

Date: 26-05 -2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Prediction Rating of Best Cuisines in Country Capitals Using Machine Learning Algorithms

Priyanka Vesireddy
X17152631

Abstract

The exciting thing for the people who loves food (foodies) what makes them happier is food wanted to see, enjoy the flavour of the dish, wants to get engaged with the new varieties of platters across global cites. As this is the inspiration from Zomato API it is the exploration for the individuals who wanted to try different varieties of platters within their convincing finance. It's study also relates the individuals who are looking for the point that has wide range of outlets which supplies different varieties of platters which they exactly desire. This Research study suggests a platform, procedure for studying the features which are static and real-time and wanted to use some models on Zomato recommendations on the prediction 'aggregate rating' and analyse its features to show the kind of cuisine is served, customer recommendations and worked on the real time data from the Zomato API blog where the data engineers can follow procedure to study the features and use in there deployment purpose as they do not have the readily available code for their production. The basic novelty is that with the serving of proper cuisines and considering the consumer priorities personalized recommendations of dining rooms are brought for where ML techniques is used like analysing and adding the new features. The research study is worked on the data file from Zomato API blog in the configuration of raw. json folder. It uses set of methodologies and algorithms which is CRISP-DM, Feature engineering on analysis of currency from different origins and conversion, Exploratory data analysis on rating, Regression, Random forest, Decision tree, Gradient Boosting, XG Boost. The obtained results are good enough for the used procedures in this work. The overall used models give their better contribution of accuracy. The future scope is discussed.

Keywords: Rf model, testing features, EDA, XG Boost, Regression Analysis, Boosting

1 Introduction

It is always surprising things how the food supply industry is growing rapidly with the inventions of the innovative food cuisines for the foodies all over the planet. The main motivation behind the paper is that to help the group to find the proper cuisine present in

there nearby locations and predict the total ratings for the individual eating places separately. The database used for our research is extracted from the Zomato API Data Survey which is excellent search engine for the gourmets for finding best cuisines within their finance. It is also the helpful inspection for the users who wanted to justify their money for the dish they order and to satisfy their cravings from various corners. It also makes the users to know and track the different varieties of cuisines from Indian to Italian, Mexican to Thai, Chinese to Greek, French to Korean, Turkish to Spanish and much more and the quality of the eating places supply and also to find the corner that has maximum mart of eat-outs on the easy go. The Zomato API provides user to access about the raw details of about 1.5 million of restaurants across 10,000 country capitals where user gets chance to track 3 categories to search eating place by name, dish and area and other category gives the data about the ratings, location and dish from scrap. The final but not least category gives about which if the Zomato cuisine index exhibits excellent area of eat-out to visit in a city including the mapping location. Finally, this search engine app is one of the boon to foodies for keep on tracking about the latest cuisines of food which makes them easy and to get the data about the outlet details gives the allowance up to 1000 toll free service calls every 24 hours by signing from their accounts. Some of the eating places are also popular to provide some entertainment like inns, shaft, Buffets and puddings etc in the durable prices. With this there would be great competition between the start-up eatery to historic eatery. Here our Zomato information focuses on the anthropology of the mapping area. In the food supply industry the demand is greatly enhancing as everyone is turning to the foodie and showing interest to taste the different varieties of cuisines but there were also few eateries which could not expand there market and brand and still striving as most of the eateries are supplying same variety of dishes which other eating places also do which also might be the reason to change and which is also the challenging task for attracting the customers. For the people who visit the city for the first time for there vacation, work to track the location it is also necessary to check for the demography of the area. In our study we also check for how many outlets provide only the vegan food in the convenience of the customers were also concentrated. So, from the data from the no of votes given by the customer there is possibility to know about the recommendations and ratings colour about the public interest to visit again are presented in the database extracted. For the start-up eat-outs and provide their service to the people needs to know some of the points like about the anthropology of the area, varieties of dishes offered and variety of food famous in the area should be the core things to be focused.

The detailed research is discussed in the further sections. Section 1 gives the brief introduction about the practice. Section 2 gives about the literature survey which is the groundwork already done in this field of study. Section 3 gives about discussion of the overall methodologies used along with some of the machine learning techniques. Section 4 gives about the sketch requirements and planning structure used. Section 5 gives the over all applications of the research. Section 6 gives about the estimations of the algorithms considered and summarizes with briefing the idea of the work and gives the inputs for the hereafter research.

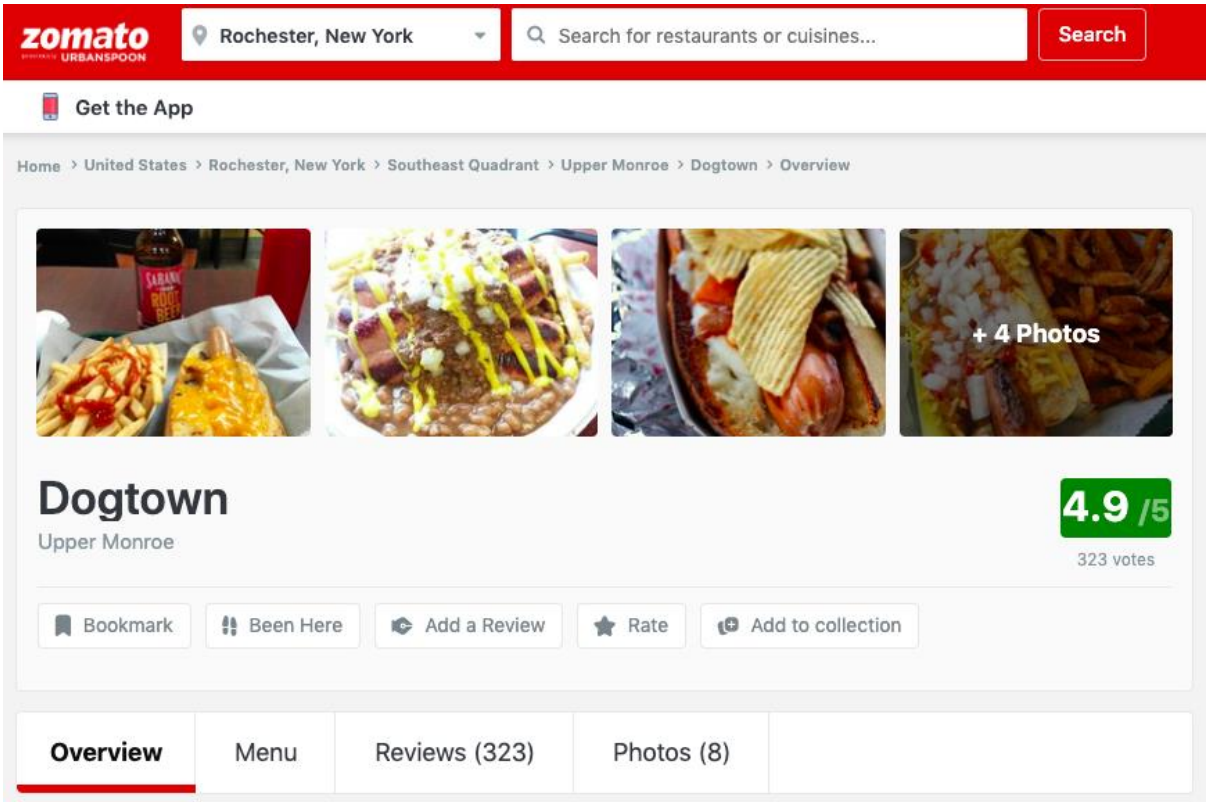


Figure 1: Zomato Search Engine

2 Related Work

An excellent conduct of the effort is already contributed in the field of the food supply service. This related survey is categorized into 2 divisions.

2.1 Literature Survey of Zomato search engine to track Best cuisines

Starting this year 2020, the author (*Lavanya, D., 2020*) food occupies major space in the trade production which makes $490 billons as per 2013 estimates and $894.98 billions is expected in 2020. To discuss about Zomato is the tool engine that brings the connection between the food supply companies, eating house and the gourmets in contact with each other. It is the active search engine tool on the social platform like twitter, Instagram. For finding best cuisine in the excellent eating house from every 19 out of 24 nations. It also provides services for dine-outs, online order, delivery services and many more. It also has enlarged its services to all over the planet. It mainly gives its full services in the field of the checking out for the better food cuisines all over the world. For retaining of the betterment Zomato has the need and scope to improve of the more applications like drive inflation, conferring, live updates of discounts for the dishes, sponsorship of new eating house, staff schooling, making ready of meal dispatching execution instalment offered meal dispatching application which also includes supply chain services and front end and back end unification can be furtherly improved here.

In the early of 2018, Zomato e-trading business also has also enhanced this is all of number of people who use web browsing has been increased up to 90 millions in the year of 2013 followed by 185 millions over the span of 2014 which has increased its sales as most of the people has turned to the hungry person and would like to tastier different varieties of dishes to increase its business across worldwide.

The author (*Sahni, N.S. and Nair, H., 2018*) wants to introduce their concept about direction function in the field of cuisine where promotion about those are exhibited. Here user approaches for the better discount and updates of the latest cuisines where the further promotion about the cuisines to be taken into the netizens through the pop-up ads which is the under direction which follows order-using planning. With this certain using of planning by many of the eating houses the finance is shown up to $2.7 million under the returns for participation of promoting their brand. Furtherly for this direction results were focused on the manifesto plan and there was also the financial crisis.

The author brief's (*Kulkarni, S.S., 2013*) has introduced and used the strategy of IOS approach which gives about the proposal of integrated with Watson API's which was victory of launching the main use of this approach is that it gives access for the participants about required facts in the right time they needed. Its also helpful to clarify the doubts of the shoppers where this approach ties up with the rank train replica for solving their doubts. Watson is the effective mechanism in the concept of coding language. Simply if the data given needs no transformation process then training needs to be given by Watson mechanism is given properly for the proper outcomes. It also takes the help of the NLP device that are in service. Next is the testing phase which is not done properly then the Watson mechanism fails to give proper outcomes.

The authors (*Heng, Z., Roy, S., Yap, K.H., Kot, A. and Duan, L., 2018*) tries to propose about customized understanding refined based grid which is PKD-Net for versatile nutriment identification. This grid identifies nutriments from the baking canvas seized by the android or mac devices and is constructed from the Mobile Net planning which assembles acceptable for upcoming stationing on Android or Mac instruments. The PKD-Net approach deserves excellent grouping execution in the neural network construction using two crafts 1) customized study of ultimate consumer nutritional inclination 2) comprehension purification of huge pedagogy grid.

2.2 Techniques used to predict ratings

The author (*YARAGATTI, P., 2019*) mainly focuses on the Zomato is the good place of action for linking so many clients with the e-cuisine manifesto. For this study the author has considered the k-mean cluster design. With the help of the design work we get chance to check about the new start-up outlets established with the help of the eye of mind pictorial representation. It is well established with the shiny web applications and gives excellent gateway proficiency of plots for Zomato particulars. So, from this we get the flexibility to

check out for the leading eating houses in the discrete conurbation and variety of platters they serve, and, in this story, it is also important about the scrutiny grades of the outlets to contrast one eating house with other eating house in particular area. It is pliable for remittance, problem solving trial, even also for the Zomato netting fidelity points.

The paper focus on the (*AM, S., 2018*) for explaining the concept of that service scape takes the half-way in moulding the clients encounters, steering buyer conduct. This concept has already manifested that generating and magnifying shopper digital value is fabricating positive encounters for the consumers such as pleasant dine-outs. Also, for making customers happier some of the eating houses tends to maintain the clean and musical eating houses for the peaceful surroundings for dine-outs. Due to following of above things there is possibility of enhancing the business with the increase of clients. So finally, quality feast gives good sales, supervisor who plan for service scape deliberately are moving to forefront the event.

In the Year of 2018 (*Almeida, J.C.E., 2018*) there was research conducted on the who are allergic to some of the nutriment so our concept to bring on the screen is to give guidance about certain type of eating house to prefer for the persons who are hypersensitive for some of nutriments which is absolutely a serious issue which can be existence alarming. The recommendation for the above issue is that they should give their appraisal about their encounter in the outlet blog which makes easier for other shoppers when they are first time visiting the outlet. The used algorithm for the concept is the thematic survey procedure. In this they clearly explained about the issues of individuals surviving with the hypersensitivity, eating out is difficulty, customers tend to find for non-hypersensitivity outlets.

In the recent inquiry author (*Stoitzner, N.M.U., 2020*) inquiry show his forecasting about proposition of random sampling is considered collect the details of all the customers who utilize wired appraisal from social posts and accounts from which by running outcome performed is sampling is the restriction. Here the offender allocated for class might be about the standard of cooking, assistance standard, outlet environment will mostly be focused. So, in our upcoming research there might be contrast between several eating houses first class feast which gives better accuracy for groundwork.

The author (*Saha, S. and Santra, A.K., 2017*) wants to interpret about the networked responses are taken from the buyers as an essential part building resolution before picking up any outcome. In this concept we take shopper content view into main consideration for grading eating houses accessible. For acquiring precision for a referrer organisation realize the customer taste. For performing this procedure first, we perform sentimental method by computing buyer view on the grub element. With this we can also review social nostalgic impact. After that we have observe for stature of the dining room from tender scattering from the bunch of the customers. The conduct appraisal of our trial outcome is established from idea of collective purification. This study can tell clearly that buyer bias can improve proposal organisation.

In the year 2017 author (*Perera, I.K.C.U. and Caldera, H.A., 2017*) investigates to predict the customer ranking by opinion excavation amalgam of data processing and robotics. Using Opinion excavation, we can inspect a form and its final plan can be pulled out. Above cramming have pre-owned certificate extent and ruling extent opinion excavation to settle the subject. To withdraw from the form POS label colony, interpret and Senti WordNet has been pre-owned. For searching the opinion term in our further investigation can be bitterly upgraded. As per the suggested plan beam, feelings have been detached during the digest phase. They might give powerful belief commitment manufacturing operation. In the code string if they are operated then real sight of individual opinion can be pulled out. Senti WordNet is realm glossary. If additional number of Senti WordNet folders are generated rectified opinions can be recognized.

As per idea of author (*Sharma, S. and Singla, A., 2018*) in related to the outcome shown for any sort of assistance given, networked evaluation has suited in the accessible form to allocate an individual buyer encounter to the other buyers. In this scenario the appraisals are the effortless ways to study about aid of the organisation for the buyers. In that group Zomato has its place used by plenty of users for commanding for nutriment in the wired form and cohering about the appraisal about the eating house. For this we have grouped all the eating places into several different groups depending upon there assistance variables. For this machine learning methods like decision tree and random forest have been used. From the outcome it has noted that decision tree classifier has given the precision around 63.5% and for the random forest method it is given precision around 56%.

In the year of 2016 the authors (*Chandankhede, C., Devle, P., Waskar, A., Chopdekar, N. and Patil, S., 2016*) has set a experiment that to consider the appraisals of the shoppers about the eating houses and there assistances. We have achieved victory in constructing an organisation which we mainly focus on the feature extent appraisals by backing implicit method. The first step to pre-treat the pulled-out appraisals with the help of WordNet, NLP library functions. By performing this step, we have noticed that precision of the organisation has come over by use for implicit method when compared with the explicit method. For finer comprehension of standard of the eating houses we have given our outcomes to the shoppers in analytical form.

The paper mainly focuses on (*Hegde, S., Satyappanavar, S. and Setty, S., 2017*) challenges to open the eating house trade by suggesting eatery trade structure. For this exploratory inspection we thought that bank card would suitable feature and Monday from the details of data to be packed day with people for the season. From this experiment the output raised is to 10 adjacent eating houses acknowledge that area might also the reason to upshot trade and offer the assistance and conveniences that are casually expected. This foundation can efficiently utilize to enhance the production of eating house which raises the increase of sales and trade.

The paper aims (*Hegde, S.B., Satyappanavar, S. and Setty, S., 2018*) that peak dishes requested to serve by many of the shoppers. So based on this we have suggest nutriment grouping structure. As discussed two groups would be four-route snack grouping s the first

group we divided appraisals into constructive and destructive sentiments with the help of SVM classifier and in the second group we have appraisals into four routes apparently, jumble, greens, platter and pudding with the help of same SVM classifier from this we recognized the peak dishes mostly ordered by the shoppers in this entire routes.

The author (*Goel, M., Agarwal, A., Thukral, D. and Chakraborty, T., 2019*) narrates his work on outlining Fiducia, a guiding organisation which recommends the eating places as per the wish of the consumer nutriment piece inclination depending on the appraisal given by the customers on the Zomato blog concerning of the eateries. Here appetizers are also important in our consideration that are served in the eating places as they also play a key role for the customer satisfaction which suggesting the best eating places on web.

The author (*Modak, K.C. and Sinha, K.*) experimented regression method on the data, the outcome obtained is from 8 independent variables 4 variables are obstructively connected to the happiness of the shoppers for networked nutriment dispatch assistance. Finer reduction is one of the keys here which gives happiness to lot of shoppers when contrasted with others. Fine reductions, fine selection, nutriment standard and remittance type are constructively connected to the shopper happiness. So from this overall conduct it is proposed that wired dispatchers has possibility to give fine reductions which is essential variable for shopper happiness.

The paper (*Anjum, R. and Dev, H., 2016*) frames about that the contribution done is with help of term transformation in all types of the events which does not indicate 100% precision. But in some situations, uncertainty may arise due to the organisation code is content based. Senti word cannot be correct in most of the affairs although it has above line precision so always radical term should be appealed. Terms from the appraisal bench should be correctly coated. Finally, we are looking to grow the productivity of classifier helped.

This paper aims (*Verma, A. and Biyani, P., 2017*) about appraisals acquired for the eatery from the Zomato blog depends on the several characteristics. It do not mention the full story about the eating places such as excellent nutriment of the eating places and their atmosphere and so on. In our inspection our major task would be not going through all the appraisals just making things easier by responding to all the queries with the help of word2vec.

The author (*Chowdhary, N.S. and Pandit, A.A., 2018*) has done the fantastic job by using classification method dummy appraisals is the misleading to certain customers and is also menace to the e-food trade. It is fraudulent activity but is still suffered in the wired food platform. It suggests two category of attributes which are customer appraisal number and word number attribute. In this process the researcher has used random forest and naïve Bayes algorithms to check for authentic and dummy appraisals. Naïve Bayes method helps to identify authentic appraisals. It improves the correctness of the model along with the Naïve Bayes classifier. It is the hours absorbing operation parallelly checks for dummy appraisal by using supervised techniques.

3 Research Methodology

As our research conduct runs with the help of data mining process mainly includes collecting information from different set of databases and resources which involves in the replica construction and used for resolution building. Data mining is the amalgam of some mining methodologies in which are three types KDD (Knowledge Discovery & Data Mining), CRISP-DM (Cross-Industry Standard Process for Data- Mining), SEMMA (Sample, Explore, Modify, Model, Assess).

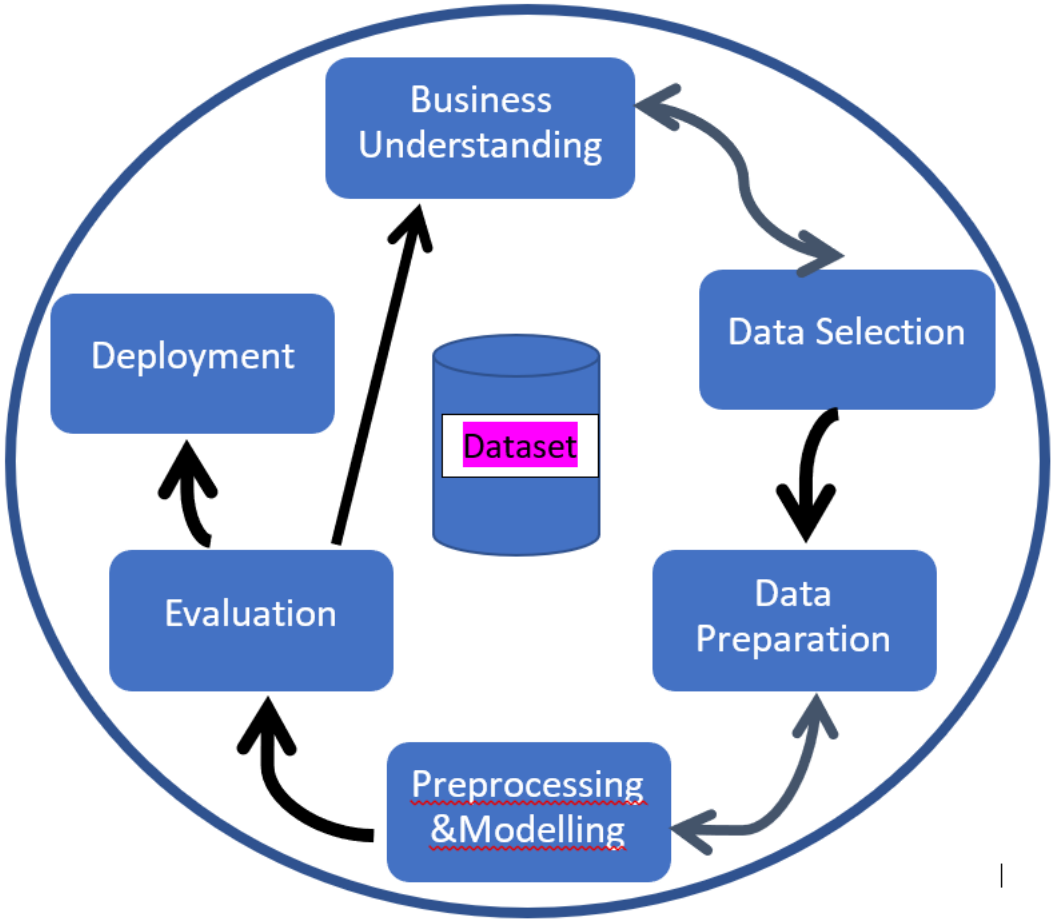


Figure 2: Methodological steps

For performing this we have considered the Cross-Industry Process for Data Mining (CRISP-DM). Our experiment distributes about to track the best cuisines from dining rooms / prediction of customers rating for the dining room. The overall methodology involved is sketched in the above diagram.

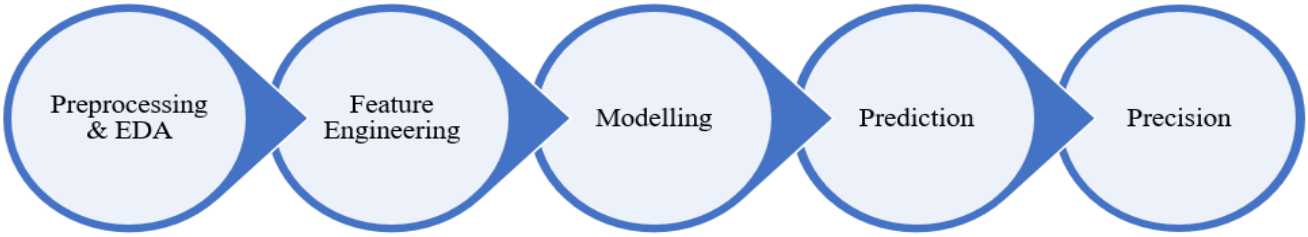


Figure 3: Step-by-step procedure of preprocessing

I Data Fetching: The loaded data file from Zomato API blog in the configuration of raw. json folder has overall 9552 rows in total. This collected data has been put down separately in the csv format folder. Here we are running test to forecast one of the attributes from the dataset which is 'Aggregate rating' which are constructed from rest of attributes. Initially data set is picked up from the Zomato API site which is in the structure of raw. Jason file. After this then the gathered data file is preserved in the independent .csv file.

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Lon
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121

Figure 4: Prediction of attributes based on the 'Aggregate Rating'

II Data Transformation: After the data file is picked then when we observe the file, we come to know that there were text columns, categorical columns and numerical columns. In this transformation phase we try to report about the remaining attributes present in the data file. Here the main logic would be there were total number of counts is that 21 columns present in the data file from which we do not approach for help of text column for running our test Model. Here we also inspect total number of entries present 9551.

	Restaurant ID	Country Code	Longitude	Latitude	Average Cost for two
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000

Figure 5: Transformation of Data

III Checking Null values: It is not easy to get the desired outcomes and plots with the presence of null values in the datafile which is commonly seen like other data files. With the help of data functions, we check data file from 0 entry to 9551 entries in which it has shown no null values so hence the dataset is clean to run the approaches.

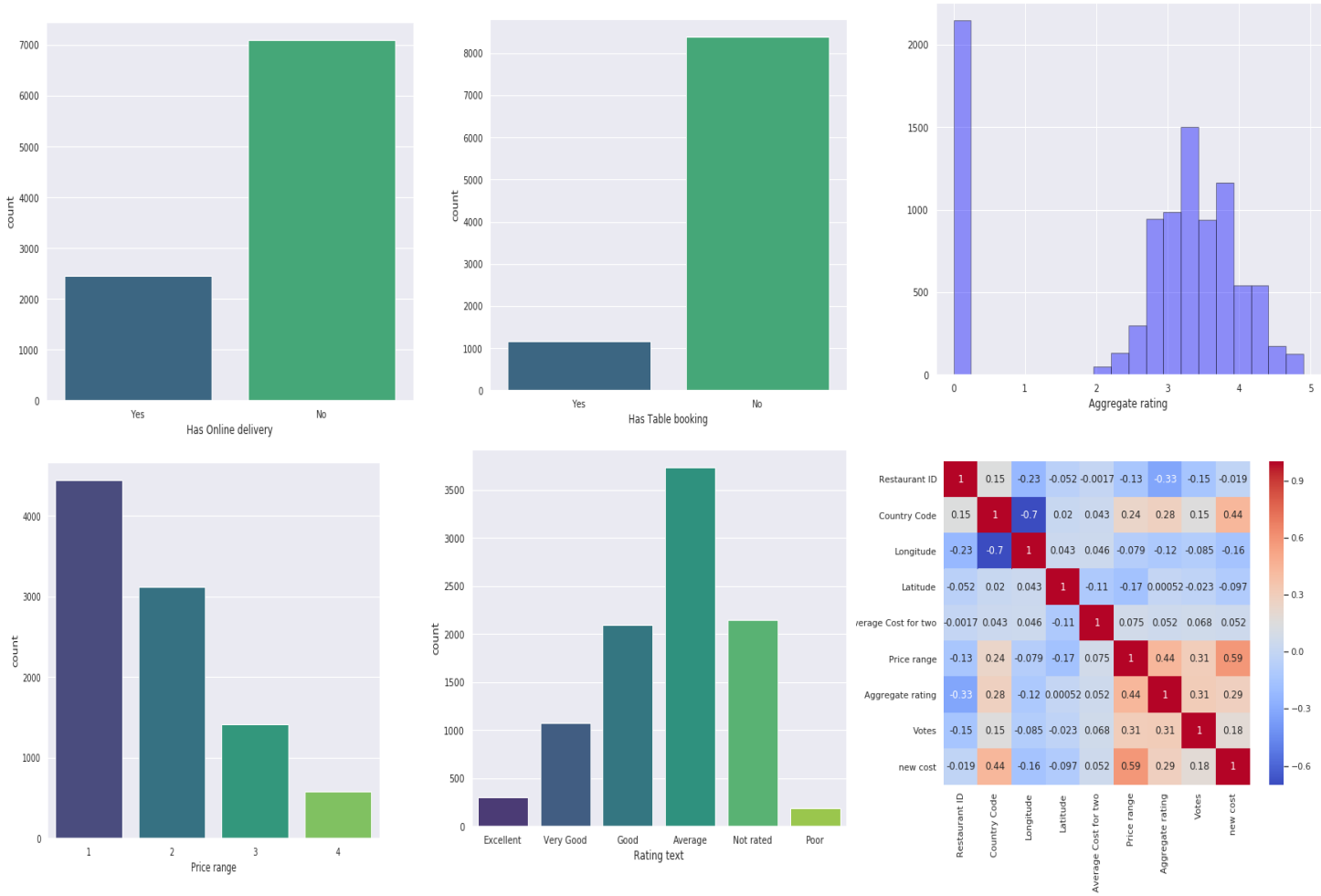


Figure 6: Data visualizations for better interpreting of data and Correlation Matrix

IV Retrieving Statistics about Target Attribute: Here we have reached to get the outcome for mean of ‘Aggregate Rating’ to be 2.66 and the standard deviation noted to the value to be 1.51 including bottom score to be 0 and the topmost score to be 4.9.

V Visualisations: The next involved is interpret the of data just by the illustrating of some of the visualisations on the set of attributes listed from the dataset. The first illustration of the plot is based on the attribute ‘Has Table booking’ which is the plot between counts and table booked which preferred not booking table before they visit an eating house. Next two illustrations is about on the two attributes ‘Has Online delivery’, ‘Is delivering now’ which shows the visual images between counts and has online delivery, is delivering currently which makes us understand that it seems to be zero deliveries. If we investigate the count of dispatching includes below.

```
No     9517
Yes      34
Name: Is delivering now, dtype: int64
```

Figure 7: Count of Present Deliveries

We can sight from above outcome that very a smaller number of restaurants are currently involved in deliveries. Because of occurred variance we cannot utilize this attribute. If we want to use another attribute ‘Switch to order menu’ but as per the outcome obtained with one only one group, we will not utilize this column.

```
No    9551
Name: Switch to order menu, dtype: int64
```

Figure 8: No group for Switching to order Menu

Another chart that we try to demonstrate is based on the ‘Price range’ shown above is about the list of prices for all the cuisines in which we can see all the cuisines are served within the customer budget from several restaurants. Another grid we try to show is based on the other attribute ‘Rating text’ which gives us the normal scattering, so we have chance to utilize this column. Next grid is associate to ‘Aggregate rating’, we can see that there still normal scattering but there are few customers who indicated their appraisal to be zero on few eatery.

VI Feature Extraction: Next we utilize the step of feature extraction to experiment the dataset in every possible way for this we use the ‘Currency’ attribute as per the estimation we are provided with 12 different categories of currencies each for use of convenience we convert into the USD($). The above generated correlation matrix is all about ‘price

range', 'votes' and 'new cost' with gives exact mark of correlation expected on those attributes so will make use of the correlation.

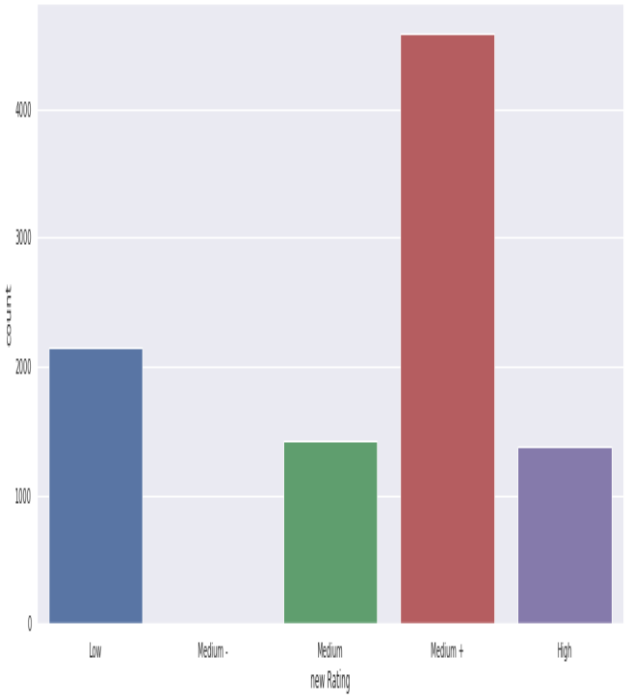
VII Exploratory Analysis: This step is last but not the least step it is mainly performed to find out the relation between attributes with each other and their targets. In this step we try to attach another characteristic for better experimenting which is 'new rating'.

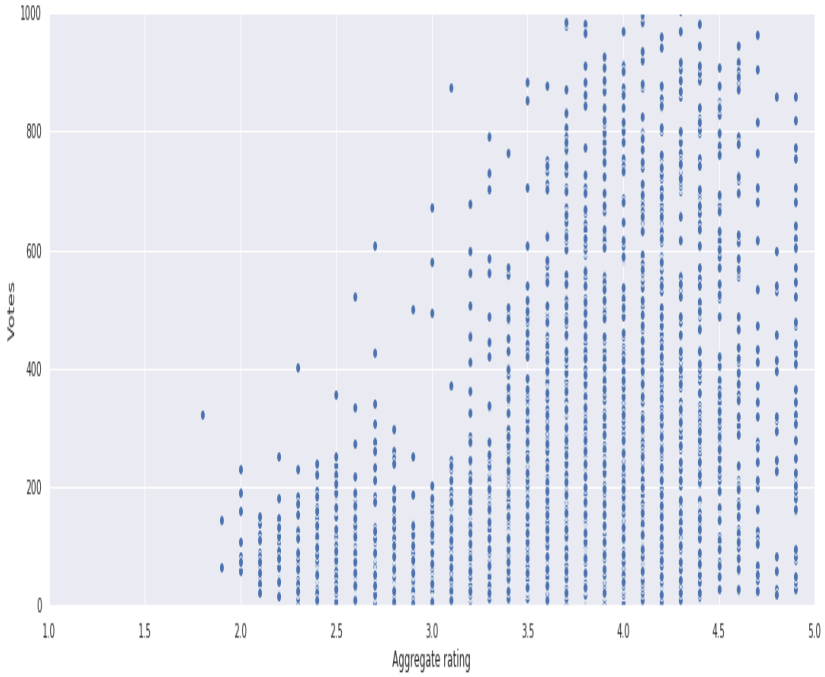
```
mask1 = (df['Aggregate rating'] < 1)
mask2 = (df['Aggregate rating'] >= 1) & (df['Aggregate rating'] <
2)
mask3 = (df['Aggregate rating'] >= 2) &(df['Aggregate rating'] <
3)
mask4 = (df['Aggregate rating'] >= 3) & (df['Aggregate rating'] <
4)
mask5 = (df['Aggregate rating'] >= 4)

df['new Rating'] = df['new Rating'].mask(mask1, 'Low')
df['new Rating'] = df['new Rating'].mask(mask2, 'Medium -')
df['new Rating'] = df['new Rating'].mask(mask3, 'Medium')
df['new Rating'] = df['new Rating'].mask(mask4, 'Medium +')
df['new Rating'] = df['new Rating'].mask(mask5, 'High')
```

Figure 9: Adding new Features

From the obtained outcome we can confirm that most of the eateries have good appraisal from the customers between 3 to 4 which is enhancing.





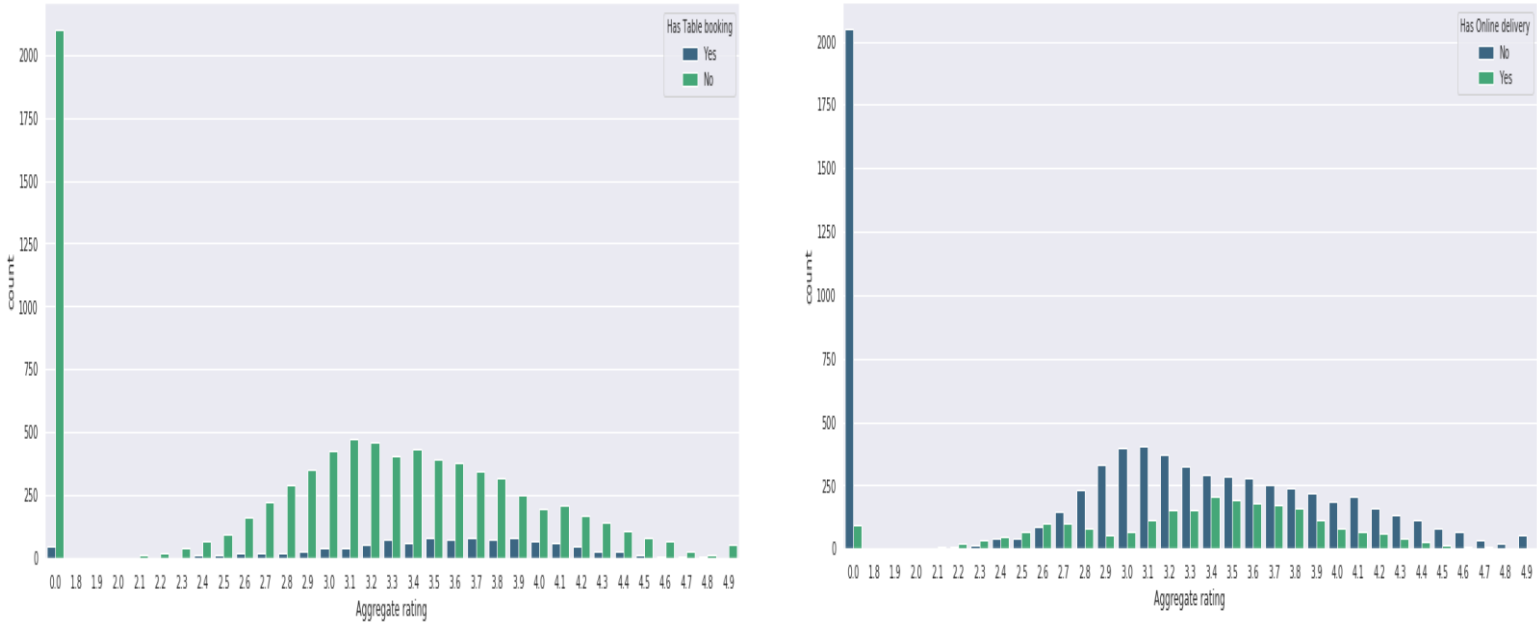


Figure 10: Exploratory Analysis for Aggregate Rating

From the above plots we have observed that there is correlation between the features of votes and aggregate rating is not well built. From the other two charts we found that immense correlation between delivery attribute when compared with the table booking attribute. We receive some aid from both above plots can further run models.

4 Design Specification

4.1 Regression

Regression is one of the machine learning models which is pre-owned for binary variable quantity. Depending on the resulting more groups of variables. When contrasted with other models it is one of the well-known models and volatile in nature. It gives nice structure of outlining from elevated arrays to regular including inside to border juncture. It has excellent continuity of variables when contradicted with the other models and 100% efficient. From this mostly non-parametric regression insane is mostly used just as to set a connection between prognostic covariate and even feedbacks. But for this nonparametric continuity of variables gives mostly borderline results and should be modified at the border juncture. Finally, when compared with set of regression model's linear regression has excellent efficiency.

4.2 Decision tree

It is one of the widely assisted empirical model for preparatory reasoning. It comparatives distinct estimate justifications that are vigorous to the turbulent statistics and effective of studying detachment assertion. It is the series of resolution shrubs from which every individual track the shrub outcome in the form of groups. Genuinely a decision tree can be observed or interpreted into a chain to enhance individual coherent. Every track from base to flag communicates for a coincidence of variable trials and shrub is a detachment to these coincidences.

4.3 Random forest

It is one of the calculated methods that can within seconds can access huge amount of data folders and easy to run and perform in model evaluation. It is mentioned that there is no specific guidelines to perform this method. There is basic condition that if the launch of the shrubs apparently grows then there will not be any change in the production, but the data processing fetch will grow. Another condition also would be if the count of shrubs raises then there is no guarantee that the production also raise. If the count of shrubs becomes twice then it depends on the connection of AUC performance.

4.4 XG Boost

XG Boost is popularly known as eXtreme Gradient Boosting technique in the house machine learning which always dependent decision tree. It is the promotion for the Gradient Boosting concept where also reliable for performance. It decodes real time problems and includes tree researching conclusion. It holds up machine learning tasks such as regression, classification and ranking. It is easily expandable from where consumers can characterize there set of purposes. It is extremely powdered and works faster in the computational environment such as Windows, Linux operating systems.

4.5 kNN

It is named as k nearest neighbour which is mostly used algorithm in the advanced mining projects. It has some of functions such as effortless performance and partitioning activities. For numerous test cases it fails to be putting stable k value. For that problem it takes aid of cross-validation approach for giving dissimilar test cases dissimilar k values. It reduces the problem related to sampling method by dividing the information k-number of subdivisions with identical dimensions.

5 Implementation

For running this experiment we used the python coding language and we import seaborn library package along with matplotlib.pyplot libraries which the graph package function used in the Python for generating the graphs. Then we import our data file which is Zomato in the form of .csv file and encodes the file. Next we view into the type of columns present in the imported data file. For running this model first, we perform pre-processing and Feature Extraction for the data file is done to extract the scatter plots etc. It is followed by Exploratory analysis just to make the data file to fit into the performance models. As the supporting algorithms to this experiment we have used Regression, Decision tree, Random forest, XG Boost, Gradient Boosting approaches which has given there set of precision values for the time of model that was run.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

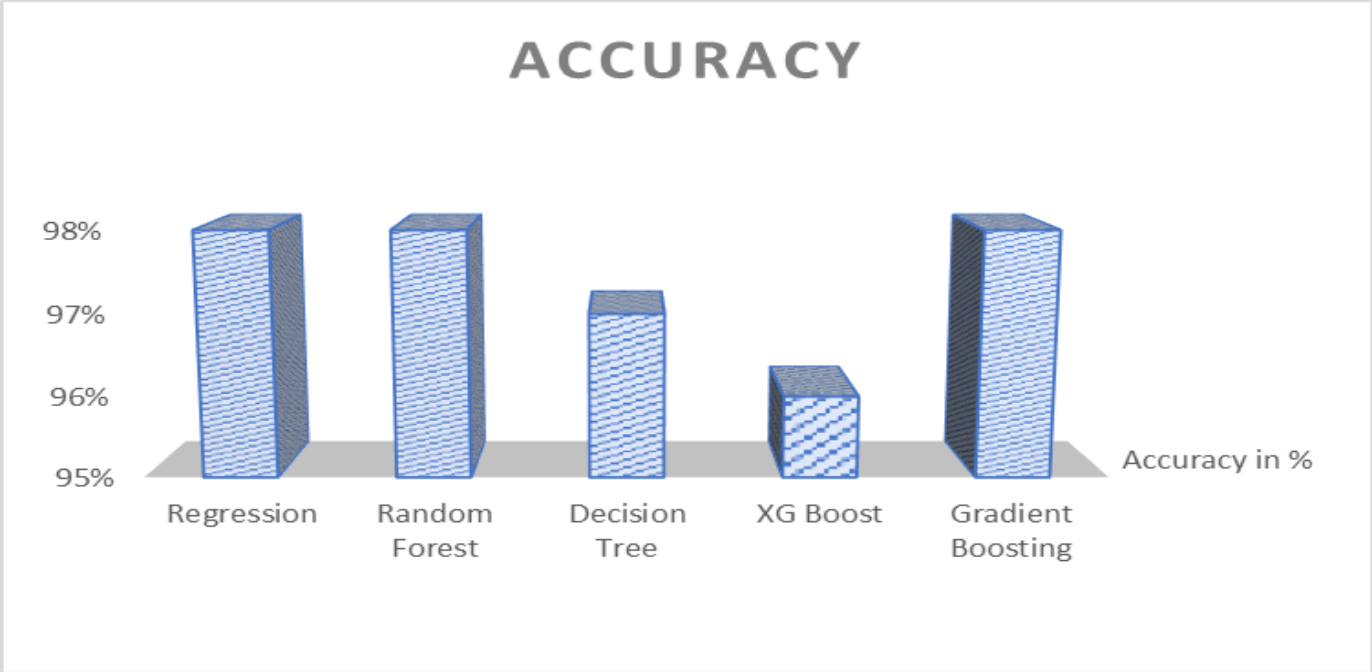
In the feature extraction stage, the currency column we encountered a different problem that different countries used different currencies so its very hard to derivate outcomes, estimations. So, we have added new feature new cost for currency column which is the unique key. By performing this step, it made easy to generate a correlation matrix. When running the model separately to predict total ratings for the column aggregate rating we have divided data into two phases training phase and testing phase for each model. When running the regression model, we have used k-fold cross validation for its stability for training phase it noted accuracy to be 96% and testing phase to be 98%. Similarly, when running the decision tree, the training accuracy is 94% and validation accuracy is 97%. For random forest model the accuracy is 98%, For gradient boosting accuracy obtained is similar 98%. So, when the R-scores are compared together it has given a normal distribution.
When running the XG Boost its accuracy is marked at 96%.

6 Evaluation

The evaluation metrics can justify by studying one of its factors called Accuracy. Here as we are tending to predict the ratings of the outlets so along with accuracy, we can also show the, error rate to check the appraisal of the customers.

6.1 Accuracy

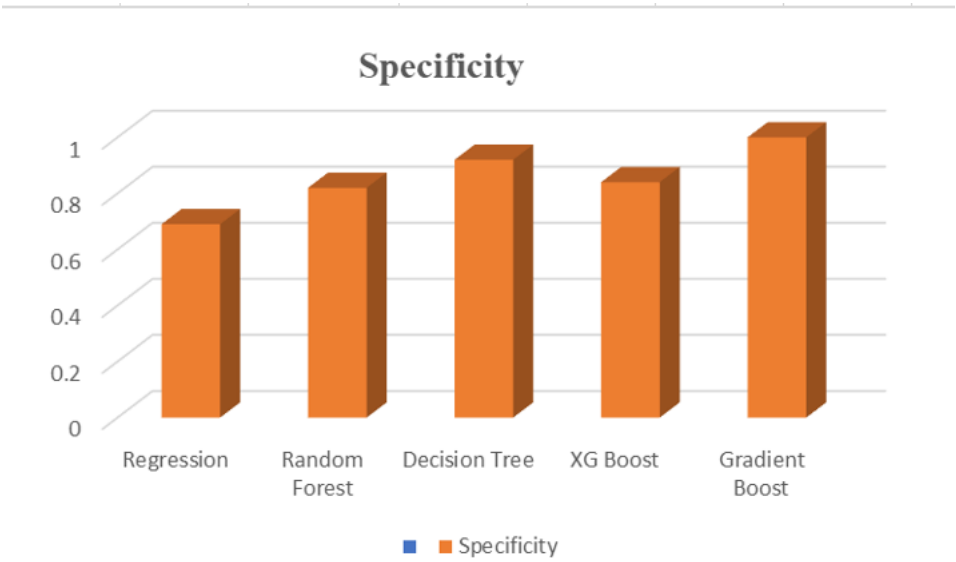
Model	Regression	Random Forest	Decision Tree	XG Boost	Gradient Boosting
Accuracy in %	**98%**	**98%**	**97%**	**96%**	**98%**



After considering the above accuracy rate performed by every induvial model has got required percentage of accuracy to run the model. If you contract one model with the other model Decision tree and XG Boost model are slightly low which is stopping the chart from building of normal different but marked as overall enough accuracy rate.

6.2 Specificity:

Model	Regression	Random Forest	Decision Tree	XG Boost	Gradient Boost
Specificity	0.69	0.82	0.92	0.84	1

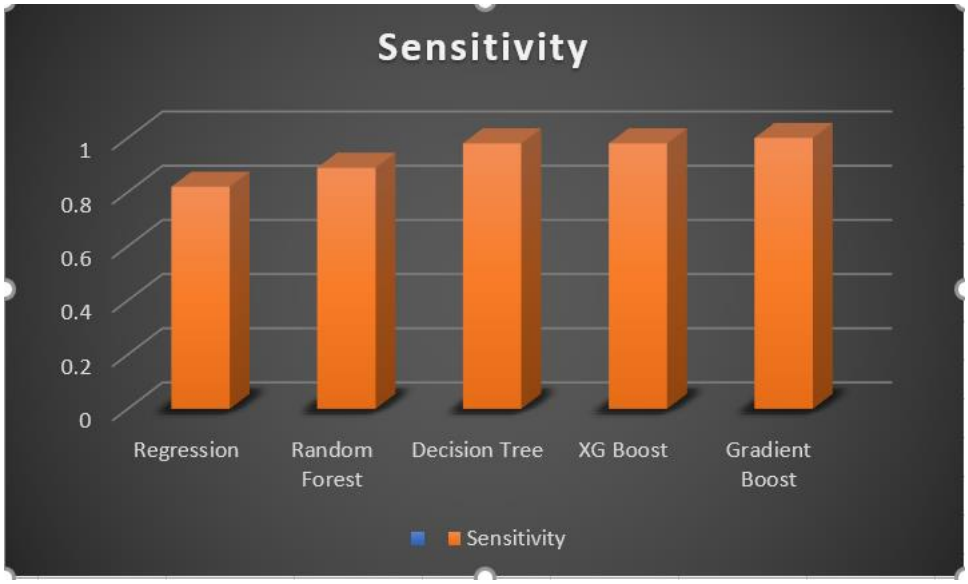


Specificity=TN/(TN+FP)

From the above illustrated bar plot for the specificity as calculated from the above formula has given that gradient boosting has highest specificity 1 when compared with the all the other models and lower is the regression model 0.69. From this we can clearly observe the distribution of the specificity.

6.3 Sensitivity:

Model	Regression	Random Forest	Decision Tree	XG Boost	Gradient Boost
Sensitivity	0.82	0.89	0.98	0.98	1

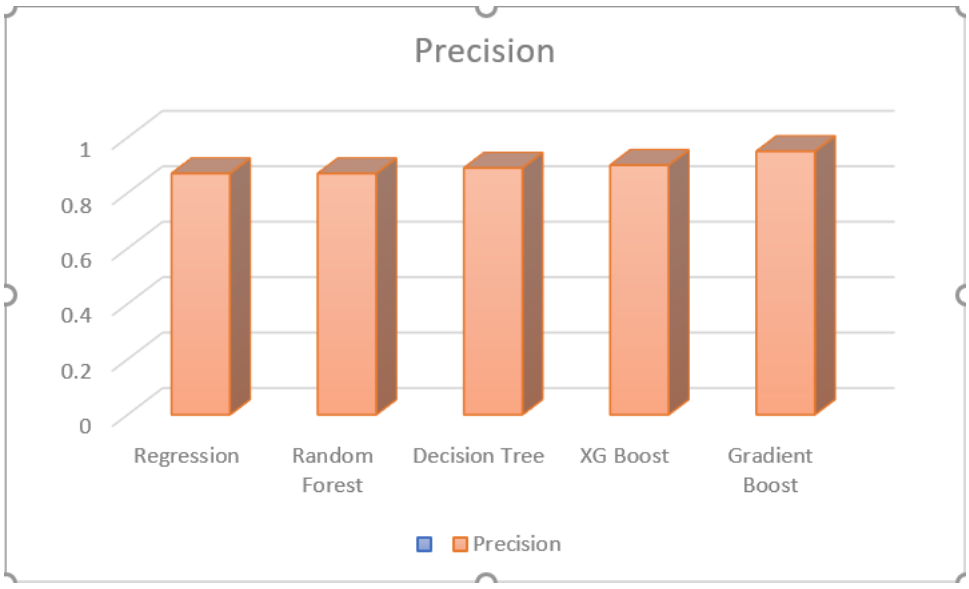


Sensitivity=TP/(TP+FN)

From the above illustrated bar plot the maximum individual recommendations for knowing about the varieties of dishes is discussed in sensitivity given that again gradient boosting has highest sensitivity 1 when compared with the all the other models and lower is the regression model with 0.82. From this we can clearly observe the distribution of the sensitivity and we can also observe that XG Boost and Decision tree has the equal distribution 0f 0.98.

6.4 Precision:

Model	Regression	Random Forest	Decision Tree	XG Boost	Gradient Boost
Precision	0.87	0.87	0.89	0.90	0.95



Precision=TP/(TP+FP)

From the above chart the individual recommendations that were recommended by the customers to try the platter is high from many of the restaurants. The highest precision rate has the value from gradient boosting with 0.95 and next place occupies by XG Boost with 0.90.

7 Conclusion

Hence, we have finally finished our experiment by performing some sort of approaches and some how gained some of the knowledge from research done. The idea of experiment is not innovative also some studies were done with same as well as different approaches on the content as part of our research we took the previous studies as an inspiration and continued to work on the idea with slight variation running few different models. From the model run we have received enough accuracy with normal distribution in their model accuracy comparison and highest accuracy could be 98%. Our idea is check ratings given by customers which would help to enhance the trade and get the proper recommendations and delivery is very low.

8 Discussion

As the dataset is pretty straight forward we did not split the data for the future research can be continued to enhance the delivery option available making users to pick as there is no necessity for them to come to the outlets the best cuisines will be delivered to their locations and also in our research we observed direct walk in for table booking outlets and low taking reservation there is scope to work even in this part more by running of different models for better accuracy. For this we need more there is also possibility to add more features to know about the opinion of customers as why they are not preferring the delivery system, not ordering online, even increase the votes and so on. There is also chance to add versions to the models and follow approach production network. We can also train the algorithms used again for better analysis of features in static and real time. There is also scoped to retrain the paradigm for the studied model and can also find a pathway to check the paradigm production.

Acknowledgement

I express my sincere gratitude to my thesis supervisor Jorge Basilio, Associate Faculty, Data Analytics, National College of Ireland, for continuous support and excellent guidance, who have also corrected my ideas for my research work for better enhancement.

I am also thankful to my family members for there strong support and continuous motivation which helped to finish my study for the idea successfully.

References

Lavanya, D., 2020. ZOMATO–MARKETING STRATEGIES AND SCOPE FOR BUSINESS DEVELOPMENT IN FOOD INDUSTRY. *Studies in Indian Place Names*, *40*(20), pp.647-656.

Sahni, N.S. and Nair, H., 2018. Does advertising serve as a signal? evidence from field experiments in mobile search.

Kulkarni, S.S., 2013. "Where should I eat?": IOS Application using Watson API's provided by Bluemix.

Heng, Z., Roy, S., Yap, K.H., Kot, A. and Duan, L., 2018. Personalized Knowledge Distillation-based Mobile Food Recognition. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)* (pp. 22-28). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

YARAGATTI, P., 2019. EXPLORATORY DATA ANALYSIS AND VISUALIZATION ON ZOMATO RESTAURANTS DATA. *Journal of the Gujarat Research Society*, *21*(14s), pp.261-273.

AM, S., 2018. Impact of Servicescape on Customer Perceived Value in Fine Dining Restaurants. *Amity Business Review*, *19*(1).

Almeida, J.C.E., 2018. Analysis, design, and evaluation of a restaurant recommendation platform for people with food allergies.

Stoitzner, N.M.U., 2020. *The effect of different types of online reviews on Instagram regarding users' choice of a restaurant* (Doctoral dissertation).

Saha, S. and Santra, A.K., 2017, August. Restaurant rating based on textual feedback. In *2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)* (pp. 1-5). IEEE.

Perera, I.K.C.U. and Caldera, H.A., 2017, September. Aspect based opinion mining on restaurant reviews. In *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)* (pp. 542-546). IEEE.

Sharma, S. and Singla, A., 2018, December. A Study of Tree Based Machine Learning Techniques for Restaurant Reviews. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1-4). IEEE.

Chandankhede, C., Devle, P., Waskar, A., Chopdekar, N. and Patil, S., 2016, December. ISAR: Implicit sentiment analysis of user reviews. In *2016 International Conference on Computing, Analytics and Security Trends (CAST)* (pp. 357-361). IEEE.

Hegde, S., Satyappanavar, S. and Setty, S., 2017, September. Restaurant setup business analysis using yelp dataset. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 2342-2348). IEEE.

Hegde, S.B., Satyappanavar, S. and Setty, S., 2018, September. Sentiment based Food Classification for Restaurant Business. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1455-1462). IEEE.
Goel, M., Agarwal, A., Thukral, D. and Chakraborty, T., 2019. Fiducia: A Personalized Food Recommender System for Zomato. *arXiv preprint arXiv:1903.10117*.

Modak, K.C. and Sinha, K., A Logistic Regression Model of Customer Satisfaction for Online Food Delivery Services.

Anjum, R. and Dev, H., 2016. Application of text mining to build a recommendation system for restaurants and their dishes. *International Journal of Computer Applications*, *155*(4).

Verma, A. and Biyani, P., 2017. Word2vec for reviews.

Chowdhary, N.S. and Pandit, A.A., 2018. Fake review detection using classification. *International Journal of Computer Applications*, *180*(50), pp.16-21.