Smart Bridge Internship

Team Name : Tech Army

Team Leader : Gopala Srilaxmi

Team members : Botla Bhavana

Pasunuri Jhansi

Aidulapuram Ranjith

Zomato Customer Review Prediction

Introduction :

In today’s digitized modern world , popularity of food apps is increasing due to its functionality to view, book and order for food by a few clicks on the phone for their favourite restaurant or cafes, by surveying the user ratings and reviews of the previously visited customers. Zomato provides a secular part where user can rate their experience of the visited restaurant or cafe.

Zomato also provides columns for writing classified user reviews. Such sort of substance provided by web is named as client produced content.

Client created content contains a great deal of significant and essential data about the food items and restaurant administrations .Since there is no control on the nature of this substance on the web and thus, these elevate fraudsters to compose counterfeit surveys to defame the restaurant administrations, to provide misguiding reviews , to generate irrelevant content regardless of the product or service, to advertise unrelated content, etc.

These phony surveys anticipate clients and associations achieving genuine decisions about the product, services, and amenities of the restaurants or cafes . In this case, Review Analysis has become vital to generate authenticated and unbiased reviews which help in avoiding fraudulent activities used to promote business by publishing fake reviews.

Objectives of Research :

This study aims to design and construct an “Online Ordering Review System”,

• To find out the customers perceptions and knowledge of Electronic food ordering that influences their buying decisions.

• To analyse which restaurants is best at service.

•The online reviews on different restaurants will have a great impact on business profit as customers will look for the review and rating online before dining in a restaurant .Thus customer review play’s an important role .

•From the dataset the customer rating can be classified using multiple other parameters.

•Based on the customer ratings and reviews profits can be predicted.

Problem Statement :

To predict zomato customer review based on some input parameters given.

Review of Literature :

In today’s digital world , food app like Zomato is widely used because it provides a platform for people to share their opinion about the restaurants and cafes they have visited . This project includes prediction of client ratings and reviews In Zomato utilizing classification algorithms . In order to train the model we make use of the dataset with previous rating’s and review’s . This project can not only shed lights on what customers value the most about a restaurant, but also provide suggestions on what feature combinations one should choose when opening a new restaurant, and how likely this restaurant can succeed .

Data Collection :

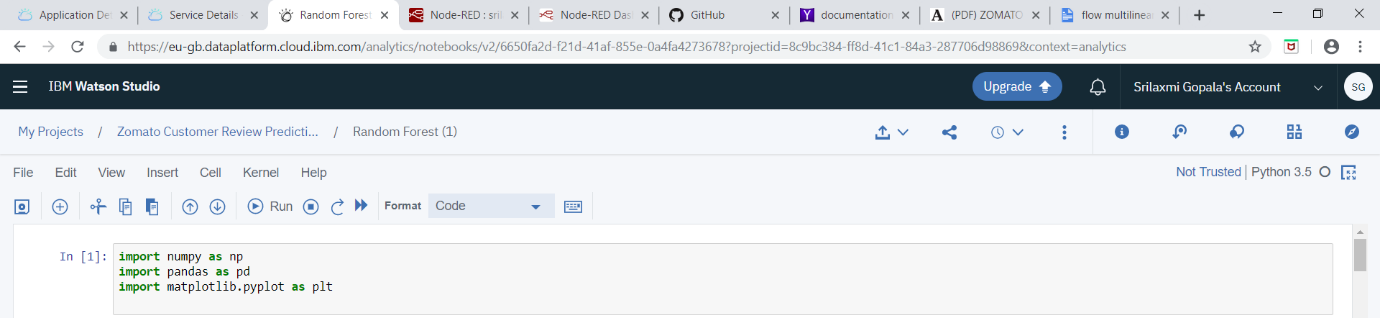
The dataset that has been used for prediction of Zomato customer review consists of 9 attributes and 9950 instances.

• The attributes present in the dataset are Average-Cost-for-two , Table-booking , Online-delivery , delivering-now , Price-range, Aggregate-rating , Rating-color , Rating-text , Votes .

• Before building the model pre-processing of dataset is required .

Steps involved in Data pre-processing are as follows :

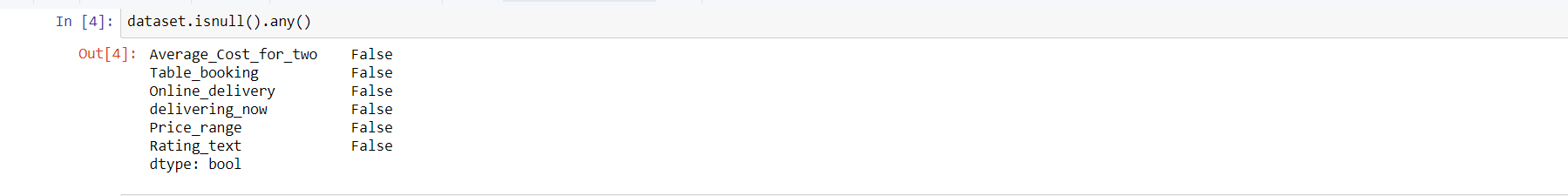
• importing libraries such as numpy,pandas,matplot



• importing dataset



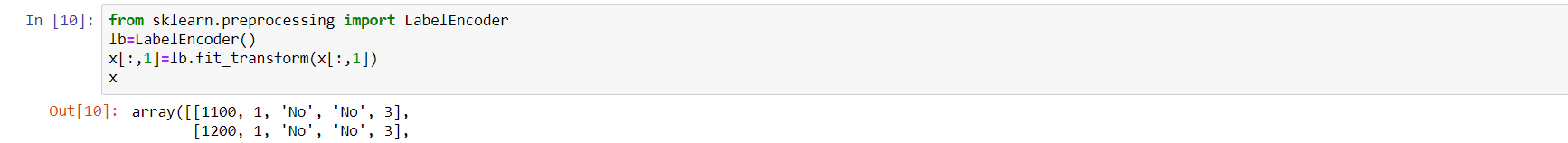
• handling missing values



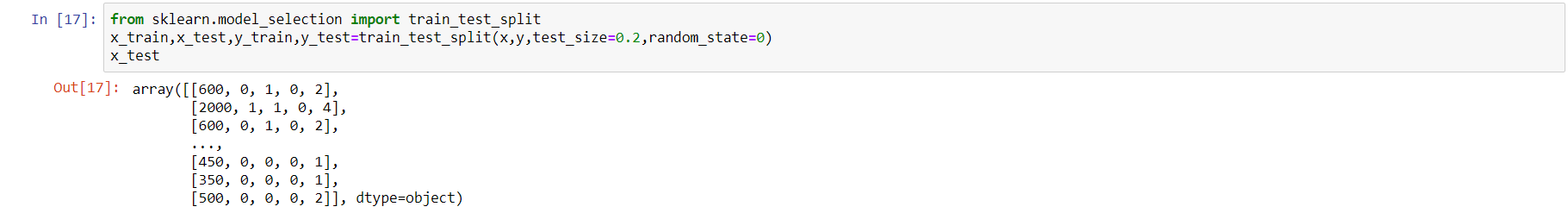
• separating independent and dependent variables



• encoding categorical text variables



• splitting train and test data .



• The missing data can be handled in three ways :

🡪Deleting the rows which are empty(missing values)

🡪Find the mean of the attribute and fill the cell with obtained mean value.

🡪Find the mode of the attribute and fill the cell with obtained mode value.

Methodology :

In machine learning , classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation.

classification algorithms : 🡪Decision Tree

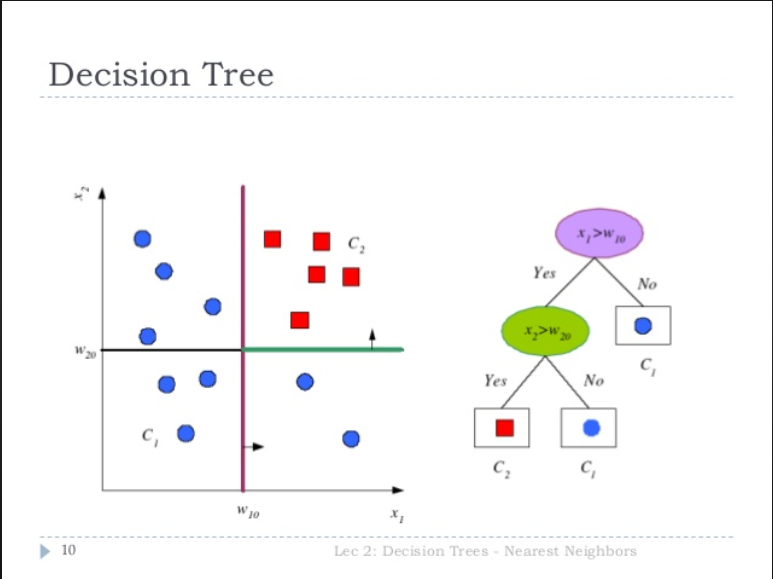
🡪Random Forest

🡪K Nearest Neighbor

## classification algorithms :

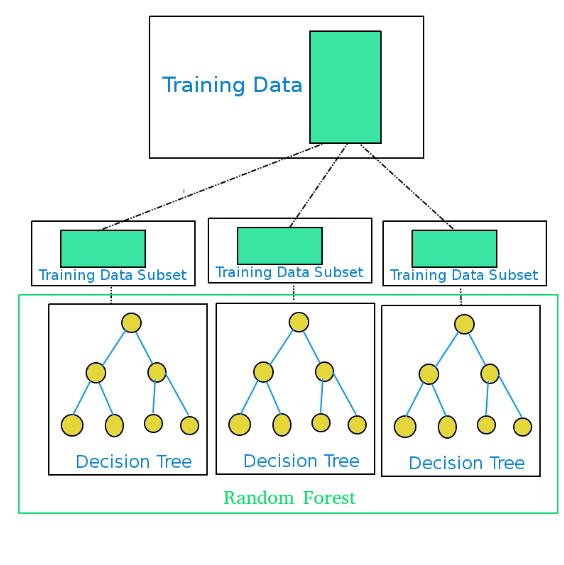
## 🡪Decision Trees:

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root  node. Decision trees can handle both categorical and numerical data.



🡪Random Forest :

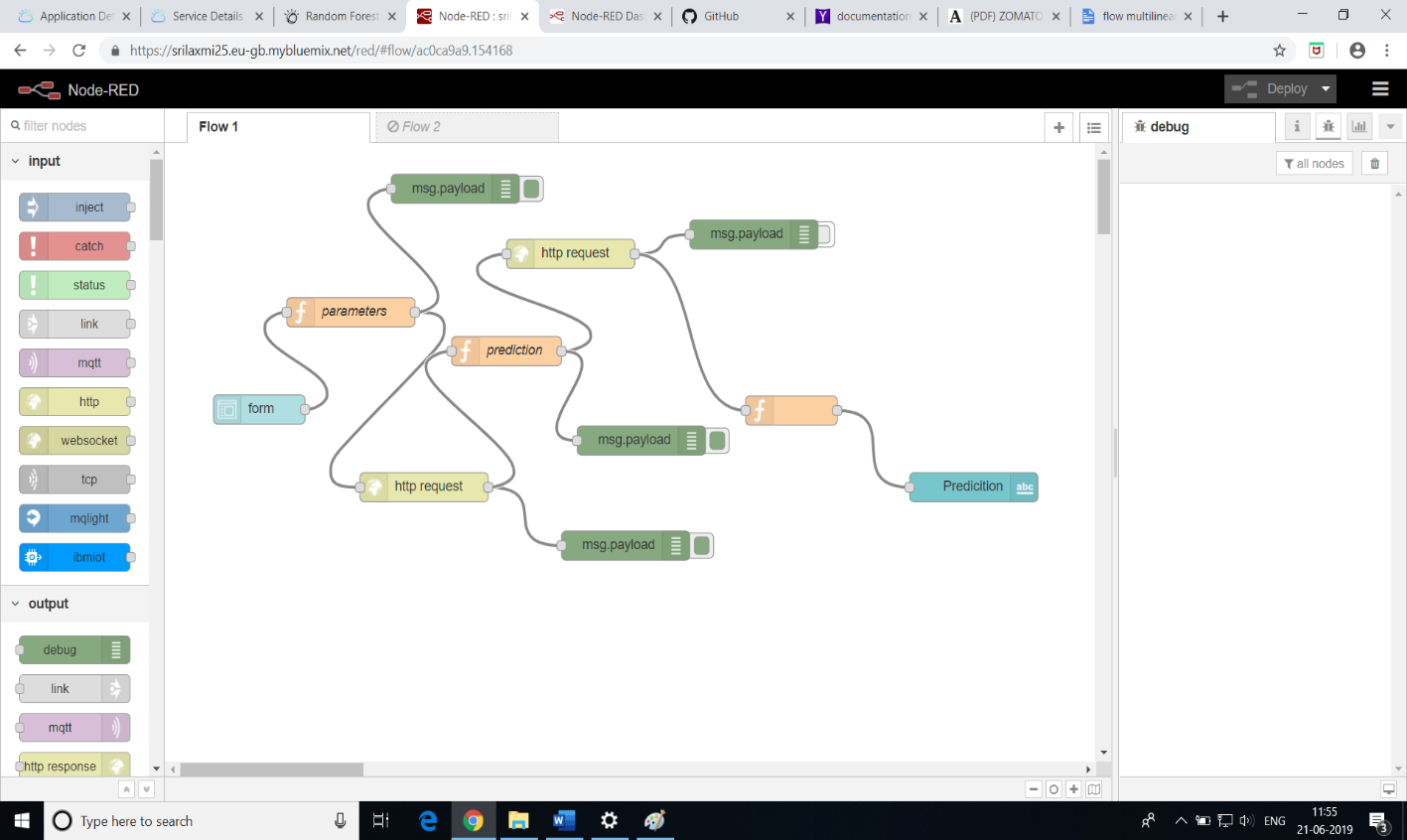
Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees’ habit of over fitting to their training set.

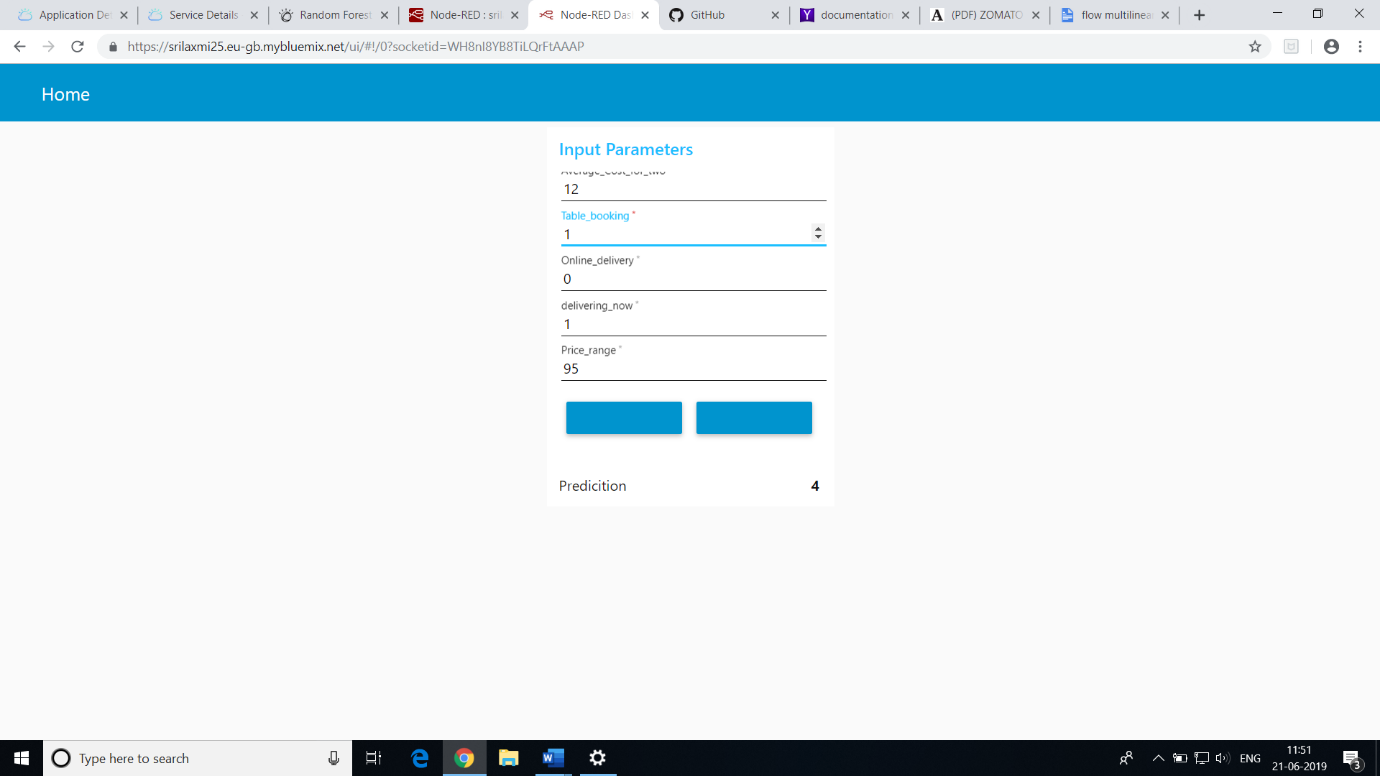


🡪K Nearest Neighbor :

 It takes a bunch of labelled points and uses them to learn how to label other points. To label a new point, it looks at the labelled points closest to that new point (those are its nearest neighbors), and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point (the “k” is the number of neighbors it checks).

Figures and Tables :





Statistical Techniques :

In classification model accuracy\_score is used to calculate the accuracy of the model .



Data modelling using Supervised ML techniques :

Supervised ml techniques in our model are

Model built using Decision Tree:

🡪importing DecisionTreeClassifier package

🡪Fitting model using x\_train and y\_train

🡪predicting values of y using x\_test

🡪calculating accuracy\_score

Model built using Random Forest:

🡪importing RandomForestClassifier package

🡪Fitting model using x\_train and y\_train

🡪predicting values of y using x\_test

🡪calculating accuracy\_score

Model built using K Nearest Neighbor:

🡪importing KNeighborsClassifier package

🡪Fitting model using x\_train and y\_train

🡪predicting values of y using x\_test ,calculate accuracy\_score

Accuracy score obtained for the model using all classification algorithms:

|  |  |
| --- | --- |
| Name of Algorithm | Accuracy\_Score |
| Decision Tree | 0.42124542124542125 |
| Random Forest | 0.423861852433281 |
| KNearestNeighbor | 0.39508110936682367 |

Findings and Suggestions :

Findings :

From the above Classification Algorithms , we have calculated the accuracy score by using Decision Tree , Random Forest , KNearestNeighbor Algorithm.

Out of these Random Forest Algorithm is slightly better .

Suggestions:

Customer Review System should suggest the policy according to the individual requirements.

Conclusion:

In this project the concepts of machine learning have been implemented for building a prediction model which can help in predicting the reviews and ratings given by the customer by training the model using supervised learning technique . This model is validated using Decision Tree , Random Forest , KNearestNeighbor algorithms .

By comparing different satistics derived from these different models mentioned Random Forest was found to be better for making predictions .