International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
ISSN: 1475-7192

Food Culture Analysis in Bengaluru

Dr.V. Asha and Heena Gupta

Abstract--- *Bangalore is a city well known for its culture and food. There are nearly ten thousand restaurants across the entire city of Bangalore. These restaurants sometimes charge exorbitantly despite poor reviews. As per the data captured from Zomato, a food ordering app, we studied the factors like location, ratings, menu diversity, etc, majorly affecting a restaurant. The price to be charged was then predicted to realize overpricing or underpricing by comparing with the actual price. The prediction problem is very important to assess the prices and preferences among people.*

Keywords--- *Restaurant, Prediction, Restaurants, XGBoost.*

I. INTRODUCTION

Hotels and restaurants act as a major attraction to any city. Bangalore also boasts of several such restaurants. Several people make use of such services to try different cuisines and dishes. These restaurants charge prices as per their menu diversity, location, delivery options, table booking, type of restaurants, cuisines offered, overall price for two adults. Thus there are several factors affecting the prices charged. The study involves numeric and categorical data. Categorical data values are not taken into consideration by any machine learning algorithm without encoding. There are different kinds of encoding like label encoding, count encoding, one hot encoding, etc. In other words, vector representations of categorical features are needed.

Objectives

i) To understand the impact of each different factor affecting the price charged by a restaurant in an overall manner.
ii) To improve data preprocessing techniques for statistical analysis for such a problem.
iii) To rightly encode the data.
iv) To employ right feature engineering techniques.
v) To identify an algorithm that provides the highest accuracy for data of this kind.

II. LITERATURE SURVEY

Work related to prediction and using prediction algorithms for various applications has been an active research topic. Various papers discuss different schemes, ensemble techniques or new approaches to data preprocessing to enhance the accuracy in prediction and to minimize the loss. The paper [2] discusses various prediction algorithms like Random Forests, K-Nearest-Neighbour and Extreme Gradient Boosting. Various techniques are compared to reveal useful insights.

The paper [3] discusses an ensemble technique amalgamating various models long short term memory (LSTM), gated recurrent unit (GRU) and extreme gradient boosting (XGBOOST), thus showing a better accuracy over other

Dr.V. Asha, Department of MCA, New Horizon College of Engineering, Bangalore, India. E-mail: asha.gurudath@gmail.com
Heena Gupta, Department of Computer Science, Mount Carmel College, Bangalore, India. E-mail: heenag2248@gmail.com

DOI: 10.37200/IJPR/V24I5/PR2020698
Received: 16 Apr 2020 | Revised: 22 Apr 2020 | Accepted: 12 May 2020

models. The paper [4] discusses how XGBoost is efficient in building a recommender system for online shopping. The system predicts user preferences during an online shopping experience, thus presenting the apt item to the customer. The paper [8] discusses a method based on Box-Cox transformation for data cleaning process. The paper [9] is also an application to predict the crude oil price for planning to meet the electricity demands.

III. TECHNICAL DETAILS

Any machine learning algorithm is to prepare a model that can be implemented on similar datasets. There are various prediction algorithms for different needs. The different prediction algorithms we used are:

Multiple Regression

This is one of the easiest and one of the most commonly used of all prediction techniques. Linear regression works with a single independent variable and multiple regression works with several independent variables. The goal is to establish a best fit line across various variables affecting the dependent variable. It follows the equation:

$y=b+w_1x_1+w_2x_2+..+w_nx_n$ where y = predicted value,

b = bias element, w = weight, and

x = independent feature

XGBoost (Extreme Gradient Boost) Algorithm

This is a decision-tree based regression algorithm using a gradient boosting framework. It is a faster algorithm compared to other forms of gradient boosting. Several trees are implemented in order to obtain the most accurate solution. Cross validation is also performed to check the model‟s validity for an independent data set.

Among the various error functions available to check accuracy, RMSE (Root Mean Squared Error) is one among them. Loss or error function is a tool to evaluate model parameter. It helps to check the difference between the actual and predicted value.

$$\text{RMSE}= \sqrt{\frac{\sum_{i=1}^{n}(Predicted_i - Actual_i)^2}{n}}$$

Gradient descent also helps to determine the right fit. The model can suffer from underfitting or overfitting. Underfitting gives the similar result for any input, so the model complexity is to be increased. Overfitting remembers the data points and does prediction, so the model complexity is to be reduced.

IV. EXPERIMENT

Dataset

The dataset consisted of several independent variables and one dependent variable.

- online_order: This tells whether online ordering is accepted by the restaurant or not
- book_table: This tells whether table booking option is available or not
- rate: This contains the overall rating of the restaurant out of 5
- votes: This contains total number of rating for the restaurant

DOI: 10.37200/IJPR/V24I5/PR2020698
Received: 16 Apr 2020 | Revised: 22 Apr 2020 | Accepted: 12 May 2020

- location: This shows the locality in which the restaurant is located
- rest_type: The restaurant can be casual, café, buffet
- cuisines: This tells the various cuisines offered by the restaurant.

The dependent variable is approx_cost, that is, the approximate cost for two people to try out food in that restaurant.

Data Preprocessing

The redundant columns or columns conveying redundant information were deleted. The columns were renamed to indicate right meaning. The null values were removed giving 43499 rows.

Basic transformations were also applied to encode data in appropriate way.

Training

The training data consisting of 43499 rows was split into training set and validation set following 80% and 20% rule. The two algorithms were applied.

a) Linear Regression: Multiple linear regression with the default parameter values was run with different encoding schemes were run.
b) XGBoost: The classifier was set with booster as „gbtree", learning rate was set to „0.3" and max_depth (maximum depth of a tree) was set to 7.

V. RESULTS AND DISCUSSION

Using the model created in the training phase, the prediction for the test data was carried out. Accuracy of 77% and 84% were achieved with regression and XGBoost. Thus if the information about a new restaurant is given, the model can rightly predict the amount to be charged. Various conclusions were drawn. The rating of the restaurant does not affect their prices. Moreover restaurants of "Dine out" category had a low rating indicating high customer demands. Location also affects the way in which restaurants have charged their prices.

VI. CONCLUSION

In conclusion, this study shows how the dataset was used to study useful insights about a restaurant"s business. These insights can thus help them to provide better food services catering to people from all walks of life. XGBoost algorithm showed better accuracy than the linear regression model for this prediction problem.

REFERENCES

[1] S. Jhaveri, I. Khedkar, Y. Kantharia and S. Jaswal, 2019. Success Prediction using Random Forest, CatBoost, XGBoost and AdaBoost for Kickstarter Campaigns. *3rd International* Conference on Computing Methodologies and Communication (ICCMC), pp. 1170-1173.
[2] X. Ma, Y. Tian, C. Luo and Y. Zhang, 2018. Predicting Future Visitors of Restaurants Using Big Data. *International Conference on Machine Learning and Cybernetics (ICMLC)*, Chengdu, 2018, pp. 269-274.
[3] U. Vanichrujee, T. Horanont, W. Pattara-atikom, T. Theeramunkong and T. Shinozaki, 2018. Taxi Demand Prediction using Ensemble Model Based on RNNs and XGBOOST. *International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (*ICESIT-ICICTES), Khon Kaen, 2018, pp. 1-6.

DOI: 10.37200/IJPR/V24I5/PR2020698
Received: 16 Apr 2020 | Revised: 22 Apr 2020 | Accepted: 12 May 2020

[4] A.L. Xu, B.J. Liu and C.Y. Gu, 2018. A Recommendation System Based on Extreme Gradient Boosting Classifier. *10th International Conference on Modelling, Identification and Control (ICMIC*), Guiyang, 2018, pp. 1-5.5.

[5] G. Cao, A. Downes, S. Khan, W. Wong and G. Xu, 2018. Taxpayer Behavior Prediction in SMS Campaigns, *5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (*BESC), Kaohsiung, Taiwan, pp. 19-23.

[6] K.D. Kankanamge, Y.R. Witharanage, C.S. Withanage, M. Hansini, D. Lakmal and U. Thayasivam, 2019. Taxi Trip Travel Time Prediction with Isolated XGBoost Regression. *Moratuwa Engineering Research Conference (MERCon)*, pp. 54-59.

[7] X. Shi, Q. Li, Y. Qi, T. Huang and J. Li, 2017. An accident prediction approach based on XGBoost, *12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nanjing, pp. 1-7.

[8] W. Qiu, 2019. Credit Risk Prediction in an Imbalanced Social Lending Environment Based on XGBoost, 5th *International Conference on Big Data and Information Analytics (BigDIA), Kunming, China,* 2019, pp. 150-156.

[9] M. Gumus and M.S. Kiran, 2017. Crude oil price forecasting using XGBoost, *International Conference on Computer Science and Engineering (UBMK), Antalya*, 2017, pp. 1100-1103.

DOI: 10.37200/IJPR/V24I5/PR2020698
Received: 16 Apr 2020 | Revised: 22 Apr 2020 | Accepted: 12 May 2020