IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 18, Issue 2 Ser. II (Mar. – Apr. 2022), PP 10-14
www.iosrjournals.org

Predicting Restaurant Ratings using Back Propagation Algorithm

Bhavsar Shachi[1], Ravi Gor[2]
[1,2]Department of Mathematics, Gujarat University, India

Abstract
Restaurants or any other shopping websites are growing their business these days. And the mass group of costumers tend to trust on ratings given to the products. Prior prediction of rating the product before launching, with the help of product features, would be beneficial for the business firm. Here restaurant ratings are predicted with the Back propagation neural network model. For experimental results three different optimizers are used. The proposed model shows 70% of the accuracy.
Keywords: *Back propagation, Artificial Neural Networks, Optimizers, Ratings*

Date of Submission: 08-03-2022 Date of Acceptance: 24-03-2022

I. Introduction

Food business is trending nowadays. The Indian restaurant industry size was valued Rs 75000 crores and it estimates to grow at an annual rate of 7% (according to smergers.com). For the good profit, restaurant ratings play an important role. Due to the pandemic situation people became more aware towards ratings of restaurant, to check they are up to the mark or not. Ratings can be in form of mouth-to-mouth publicity, newspaper or magazines critics, application or website ratings. Where reviews collected by costumers are common form of ratings.

Many key private equity and venture capital firms have invested in the restaurant industry. It is important to know the level of the restaurant for opening, expanding or continuing the business. For new launching food websites, it is very difficult task to whether include particular restaurant or not on their online platform. This study can help to analyze the costumers need and choice for restaurants. Thus, predictions can shed light on which set of restaurants to include in their websites. It also offers suggestion to bank/lenders that in future restaurant gets bankrupt or not, as higher the rating higher the people visiting, tends to less chances of bankrupting. So, predicting restaurant ratings is helpful for costumers as well as owners and business.

Machine learning models are very efficient for predicting task and giving promising results. There are many techniques available for prediction such as Linear Regression, Logistic Regression, Random Forest, Naive Bayes, Support Vector Machines, Artificial Neural Networks (ANN), etc. While working with big and abstract dataset Artificial Neural Networks gives better results. ANNs can be used for both classification and regression purpose. Here target variable is ratings, grading from poor to excellent, so it is multi-class classification model.

II. Literature Review

Huang (2012) shows solution for a service failure, as there are more chances revisiting of costumer if services are good. Multilayer perceptrons (MLPs) and Support Vector Machines neural networks were used to predict service failure recovery. For training, cross validation and testing, data was acquired from questionnaire survey. Parameters such as good or bad experience, mood, voice tone, language used, etc of the costumers were included in both models. Results showed that SVM model was more accurate than MLP. [4]

Guo et al. (2017) predicted the Toronto restaurants rating and popularity change prediction using logistic regression, Naive Bayes, Neural Network, and Support Vector Machine (SVM) and results were discussed**.** The dataset covers 12 metropolitan area of 4 countries with information of opening hours, price level, service provided, location, etc. For predictions of both the ratings and popularity change comparative results showed that logistic regression performed better than other models. [12]

Chen and Xia (2020) built Naive Bayes, Decision Tree, and Neural Network regression models to predict restaurant rankings. The models include both text and non-text data such as good/bad, environment, location, services, taste, state income, review star, etc. On training set 10-fold cross validation was performed for each model and results showed that Decision tree was the best model with 82.50% accuracy. Also, neural network model performed well with the accuracy of 82.45%. [13]

J.priya (2020) built regression models for restaurant dataset to predict ratings. Features such as name, online order, book table, rate, votes, type, etc were used for the perdition. A comparative study on Linear Regression, Random Forest, Ridge Regressor, Lasso Regressor, K-Nearest Neighbourhood, Support Vector Machine, Elastic Net Regressor, Bayesian Regressor models was carried out with the same parameters resulting random forest as a best fit model. [6]

Luo et al. (2021) used traditional machine learning and deep learning models for restaurant review rating by building a web scrapper for data collection. Also, pandemic impact on ratings and review trends were studied. As a traditional method Gradient Boosting Decision Tree (GBDT) and the Random Forest were used. And as a deep learning method Simple Embedding + Average Pooling and Bidirectional LSTM were used. Where Simple Embedding + Average Pooling showed better results in online review rating task. [14]

Li et al. (2021) made a sentimental analysis model with the help of Attention method, Bi GRU and Word2vec for online restaurants reviews. Web crawler was developed to randomly draw review samples from 130 stores. Results concluded that the F1 score of combined model was better than the normally used sentiment analysis model. [7]

III. Model Description

Artificial Neural Networks (ANNs) [8]

ANNs are excellent tools in generalizing patterns for complex models. The first neural network model (the perceptron) was developed by Rosenblatt in the late 1950s [4]. Typically organized in input, hidden and output layer, neural networks are interlinked with nodes and weighted connections. All three layers are made up of neurons which imitates the working structure of a human brain. Initially information is passed in input layer which is processed in hidden layer and prediction is generated by output layer. Activation function in neural network is used to decide which information to be passed by producing a non-linear decision boundary. In neural network the layer in which each neuron receives input from all previous layer's neuron are called dense layer. Dense layer neural network can also be termed as a fully connected layer neural network.

The parameters x_i taken as an input for model makes input layer. Then with the neurons and random weights w_i assigned to each node, values are computed $\sum x_i w_i$ and transferred to each neuron of hidden layers. Again, the computed values $\sum h_i w_i$ of hidden layer neurons h_i are transferred to output layer. The value obtained by the output layer is the prediction for model. But, at first the predicted values will contain error as the weighted values taken was random. So, to find the perfect weights to generalize the model, back propagation algorithm is used to minimize the error.

Back propagation [8]

As per the name, output values are backpropagated to generate new weights. Back propagation is a chain rule to compute the gradients where optimizers like Stochastic gradient descent, Momentum, Adam, RmsProp, etc are used to minimize the loss. The weights are subsequently updated during each epoch of back propagation and process continues till optimum weights are obtained. The general equation to update weight is $w' = w - \propto \frac{\partial L}{\partial w}$; where $\propto$ is a learning rate and $\frac{\partial L}{\partial w}$ is a gradient component of loss function. Optimizers follows any of the three procedures, either it modifies learning rate or gradient or both.

Classification and regression both can be done by ANNs. Here with the help of classification neural network, restaurant ratings in five categories from poor to excellent is predicted. The reason for considering the ANNs, is one can feed as many parameters as they want to classify n categories and ANNs gives better results in classification problems.

IV. Data Collection and Data Preparation

Here, Zomato restaurant data from various countries is taken, containing 9551 columns (from Kaggle). The given restaurant dataset includes all three categories fine dine, quick dine and casual dine. Also, dataset contains Restaurant id, name, country code, city, address, locality, longitude, latitude, cuisines, average cost for two, currency, table, online delivery, delivering now, switch to order menu, price range, aggregate rating, rating colour, rating text and votes.

The following parameters are extracted from dataset

Restaurant id, name and country code are removed as there is no need of such features in computation. Address, locality, longitude and latitude are removed as model is not based on (confined to) any specific area. As no restaurant have switch to order menu, i.e. no variation in data, so it is dropped. Also, table available, online delivery and delivering now features are dropped.

The following parameters are included form dataset

As people tend to choose restaurants which have various cuisines or different variety in menu. So, the restaurant with more choices in cuisines are preferable. Therefore, names of available cuisines given for each restaurant in dataset are converted into the number of available cuisines for respective restaurants.

Cost for food is one of the notable parameters from costumers' perspective. In dataset, average cost for two person is provided in currencies of respective countries. To standardise, all values are converted in US dollar. Also, price range and votes are included**.**

Aggregate rating, rating colour and rating text are variations of rating are related to each other in dataset. So, considering any one type of rating is sufficient. Hence, Aggregate rating, rating colour are dropped, and rating text was included. Restaurants rated by text are based on five categories excellent, very good, good, average and poor.

Table 1: Final parameters included from dataset

	No. of cuisines	**Price in dollar**	**Price range**	**Votes**	**Rating**
0	3	97.9	3	314	excellent
1	1	106.8	3	591	excellent
2	4	356.0	4	270	very good
3	2	133.5	4	365	excellent
4	2	133.5	4	229	excellent

Data cleaning

All the restaurant with the missing information is irrelevant for the model. So, columns with zero entries are removed from the data set. Also, the restaurants without any ratings are also removed from the dataset. After removing unnecessary columns, final dataset contains 8439 columns.

V. Methodology

After several experimental results, an optimal sequential model with two dense hidden layers, each having 40 neurons is developed. Same as the number of parameters in the model, 5 neurons are taken in input and 5 neurons in output layer as there are 5 different categories to classify. Model is built with the help of Keras in Python. Label encoder is used for one hot encoding the categorical values i.e excellent, very good, good, average and poor, of target parameter(rating). Where activation function 'elu' is used for hidden layers and 'softmax' for output layer. Nadam, Adam and AdaMax optimizers are used for backpropagation to select the best fit model by comparison. These optimizers are considered in model because it has ability to update learning rate as well as gradient at every step.

Adam [2]

Adaptive moment estimation is a blend of RMSprop and momentum optimizers.

$w_{t+1} = w_t - \frac{\propto V'_t}{\sqrt{S'_t + e}}$ where $V'_t = \frac{V_t}{1-\beta_1^t}$, $S'_t = \frac{S_t}{1-\beta_2^t}$, $V_t = \beta_1 V_{t-1} + (1-\beta_1)\frac{\partial L}{\partial w_t}$ and $S = \beta_2 S_{t-1} + (1-\beta_2)\left[\frac{\partial L}{\partial w_t}\right]^2$. Where V exponential moving average and S squared moving average of gradients are initially taken as 0. Default values taken as $\propto = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $e = 10^{-8}$.

AdaMax [2]

It is based on Adam using infinity norms.

$w_{t+1} = w_t - \frac{\propto V'_t}{S_t}$ where $V'_t = \frac{V_t}{1-\beta_1^t}$, $V_t = \beta_1 V_{t-1} + (1-\beta_1)\frac{\partial L}{\partial w_t}$ and $S = \max\left(\beta_2 S_t, \left|\frac{\partial L}{\partial w_t}\right|\right)$. Where V exponential moving average and S past p-norms moving average of gradients are initially taken as 0. Default values taken as $\propto = 0.001$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

Nadam [10]

It is a blend of Nesterov and Adam optimizer. Nesterov is used to update gradient one step in advance than Adam.

$w_{t+1} = w_t - \frac{\propto V'_t}{\sqrt{S'_t + e}}$ where $V'_t = \frac{V_t}{1-\beta_1^t}$, $S'_t = \frac{S_t}{1-\beta_2^t}$, $V_t = \beta_1 V_t + (1-\beta_1)\frac{\partial L}{\partial w_t}$ and $S = \beta_2 S_{t-1} + (1-$ *β2∂L∂wt2*. With V exponential moving average and S squared moving average of gradients are initially taken as 0. Default values taken as $\propto = 0.002$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $e = 10^{-7}$.

The model is trained by 80% of dataset and tested by remaining validation data in the model. Number of epochs taken is 250. Accuracy of the model is calculated by F1 score metric.

$F_1 = \frac{2 \times precision \times recall}{precision + recall}$; where $precision = \frac{True\ positive}{True\ positive + False\ positive}$ and $Recall = \frac{True\ positive}{True\ positive + False\ negative}$.

VI. Results and Discussion

Here, the models with three different optimizers are built. The obtained results are shown in the table 2. And table 3 is confusion matrix of the Nadam optimizer. The results shows that Adam and Nadam gives same F1 score, but true values from confusion matrix predicted by model shows Nadam is more accurate. Also, weighted average of precision and recall is compared. So, Nadam optimizer gives better results as compared to Adam and AdaMax.

Table 2: Classification accuracy report of proposed models

	Adam	**AdaMax**	**Nadam**
F1 score accuracy	0.70	0.69	0.70
Precision weighted average	0.69	0.68	0.70
Recall weighted average	0.70	0.69	0.70
Predicted True values	5890	5815	5904

Table 3: Confusion matrix obtained for Nadam optimizer

	Poor	**Average**	**Good**	**Very good**	**Excellent**
Poor	1054	138	43	6	0
Average	0	3253	418	44	1
Good	0	716	1132	257	1
Very good	0	117	505	447	7
Excellent	0	17	79	186	10

VII. Conclusion

Machine learning based ANN classification model is designed and optimized by suitable combination of activation functions and optimizers. The proposed model is used to predict the ratings of restaurant across some countries. The ratings are categorized into five classes from poor to excellent. The final compared results shows that model achieves good accuracy of 70% by Back propagating with Nadam optimizer. This study can be helpful in understanding the customers' requirements and in expanding the business. Further, same model is capable for including more complex information. The drawback of the model is, dataset does not include comments of the costumers which could be helpful to analyze the model giving more precise outcome. Future work can be done by applying different real life classification problems on the model.

References

[1]. Atharva Kulkarni, Divya Bhandari and Sachin Bhoite, "Restaurants Rating Prediction using Machine Learning Algorithms", International Journal of Computer Applications Technology and Research, Volume 8–Issue 09, pp. 375-378, 2019. https://ijcat.com/archieve/volume8/issue9/ijcatr08091008.pdf

[2]. Diederik P. Kingma, Jimmy Ba, "Adam: A Method for Stochastic Optimization", 3rd International Conference for Learning Representations, San Diego, 2015. https://arxiv.org/abs/1412.6980

[3]. Djork-Arne Clevert, Thomas Unterthiner and Sepp Hochreiter, "Fast and accurate Deep network learning by Exponential linear units (Elus)", ICLR conference, 2016. https://arxiv.org/abs/1511.07289

[4]. Huang, "Using Artificial Neural Networks to Predict Restaurant Industry Service Recovery", International Journal of Advancements in Computing Technology (IJACT), Volume-4, pp. 315-321, 2012. https://www.researchgate.net/publication/289463707_Using_artificial_neural_networks_to_predict_restaurant_industry_service_recovery

[5]. Hyewon Youn and Zheng Gu, "Predict US restaurant firm failures: The artificial neural network model versus logistic regression mode", 2010. https://www.jstor.org/stable/23745462

[6]. J.Priya, "Predicting Restaurant Rating using Machine Learning and comparison of Regression Models", International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), pp. 1-5, 2020. https://www.semanticscholar.org/paper/Predicting-Restaurant-Rating-using-Machine-Learning-Priya/c0e64ff351ab71e2a670a417e0ad85f89f69a6b3

[7]. Liangqiang Li, Liang Yang and Yuyang Zeng, "Improving Sentiment Classification of Restaurant Reviews with Attention-Based Bi-GRU Neural Network", pp. 1-16, 2021.

https://www.mdpi.com/2073-8994/13/8/1517
[8]. Buscema, Massimo, Back Propagation Neural Networks, Substance Use & Misuse, 33(2), pp. 233–270, 1998. https://www.researchgate.net/publication/13731614_Back_Propagation_Neural_Networks
[9]. Qiwei Gan and Yang Yu, "Restaurant Rating: Industrial Standard and Word-of-Mouth A Text Mining and Multi-dimensional Sentiment Analysis", Hawaii International Conference on System Sciences, pp. 1332-1340, 2015.
[10]. Timothy Dozat, "Incorporating Nesterov Momentum into Adam", 2015. http://cs229.stanford.edu/proj2015/054_report.pdf
[11]. Xiaochen Wang, Yanyan Shen and Yanmin Zhu, "A Data Driven Approach to Predicting Rating Scores for New Restaurants", Proceedings of Machine Learning Research, pp. 678-693, 2018.http://proceedings.mlr.press/v95/wang18c.html
[12]. Yiwen Guo, ICME, Anran Lu, ICME, and Zeyu Wang, "Predicting Restaurants' Rating and Popularity Based on Yelp Dataset", pp. 1-6,2017.
[13]. Yifan Chen and Fanzeng Xia, "Restaurants' Rating Prediction Using Yelp Dataset", International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp. 25-27, 2020. https://ieeexplore.ieee.org/document/9213704
[14]. Yi Luo and Xiaowei Xu, "Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic", International Journal of Hospitality Management, pp. 1-8, 2021. https://www.econbiz.de/Record/comparative-study-of-deep-learning-models-for-analyzing-online-restaurant-reviews-in-the-era-of-the-covid-19-pandemic-luo/10012495041

Bhavsar Shachi. "Predicting Restaurant Ratings using Back Propagation Algorithm." *IOSR Journal of Mathematics (IOSR-JM)*, 18(2), (2022): pp. 10-14.