ISSN No:-2456-2165

Restaurant Review Predictionusing Machine Learning and Neural Network

Tanbin Siddique Eidul, Md.Alim Imran, Amit Kumar Das
Computer Science and Engineering
East West University Dhaka, Bangladesh

Abstract:- Nowadays, people often judge which restaurant is good or bad by looking at the rating of the restaurant. That's why ratings are a critical factor in the restaurant business. Ratings are usually given by people judging by what kind of service restaurants are providing. So, features of restaurants play a very important role in this regard. The main goal of this research is to predict ratings of restaurant business based on features to help new entrepreneurs to set up new business. We used different machine learning algorithms like Decision tree, Support vector machine (SVM), k-nearest neighbors' algorithm (KNN), Stochastic gradient descent (SGD), Gaussian Naive Bayes. We also used a convolutional neural network (CNN) model here. It gives us an accuracy score of 97.2 25 percent which is higher than all other algorithms.

Keyword:- *machine learning algorithm, convolutional neural network (CNN).*

I. INTRODUCTION

In our daily life when we try something new or make an important decision, we requested suggestions from the community and these suggestions can heavily influence our decisions. In this current modernized world, people are more connected via the internet as a result people now often make decisions based on other people's recommendations online. Rating or Ranking plays a very important role, in almost any kind of business. These evaluations heavily influence people on making choices.

This is very much true in the case of the restaurant business. People always tend to go to a restaurant with higher ratings. Study shows that only even half a star better rating can allow restaurants have 19% more frequent chance to see out which can have a significant influence on restaurants overall business [1]. To open a new business in this highly competitive sector people need to be more careful. About 59% of new restaurants fail in their business in their opening years and about 80% of restaurants fail within the next five years [2]. Now to set up a new business location plays a vital role. A perfect location can extend the chances of success for a new restaurant incredibly. We can use machine learning algorithms on collective data, to help new entrepreneurs, what must-have features can increase their rating of the restaurant business. In recent years machine learning model improved a lot to a point that where machines are producing better accuracy levels than humans. One of the biggest examples of this is Google's inception network which suppressed human-level accuracy in image classification [3]. This implies that using the machine learning model is now more convenient than ever. We can suggest new entrepreneurs a suitable place to open up a new restaurant business. Most of the work done in this field is using obscure logic [4][5][6], where customer satisfaction is one of the main concerns.

Our goal here is to predict ratings for new restaurant businesses based on collective features. People will be benefited who are trying to set up a new restaurant and by knowing an expected rating, the business plan can be re-modified according to features.

The remainder of the paper is classified as following steps, Section II, Background study is holding the related works done on the same topic; Section III narrates the data processing; Section IV retains the description of algorithms used in this work. The analysis, working procedure, and results are in Section V and at the end conclusion, limitations, and future works.

II. BACKGROUND STUDY

There has been a lot of scientific research and work on this subject before; this section describes some of the notable works that have already taken place.

In a research paper [7] researchers tried to predict the future success of Yelp Restaurants. Here Reviews collected from the customer online are useful for predicting the future of the restaurant business. The paper indicates more about online ratings provided by consumers for restaurants and determines whether the restaurant will continue its business or not. Going through the paper, it is important to maintain a certain number of reviews and least ratings on YELP to have average to maximum customer attraction as they used the YELP dataset. It is also important to have an eye-catchy and well-crafted environment. Some other factors that affect the ratings are Food quality, employee behavior, location, etc.2 different datasets and 15 attributes were used and analyze data by categorizing them as Text Features and Non-text Features. As it was a binary classification problem, Logistic regression was used as the classifier. Accuracy was found at 67.46%. 73% of restaurants found open. The prediction was conducted based on a 1year period of the dataset.

In another work by Aillen Wang and his fellow researchers [8] tried to predict new restaurants' success and rating and find out which features controls a restaurant's success. They defined some conditions for a restaurant to be considered successful. Yelp Dataset for restaurants was used here. They performed a chi-square test and stochastic gradient descent (SGD) to find out optimal restaurant features which have the most weight. Different types of

ISSN No:-2456-2165

binary and multi-class classification algorithms like Random forest, logistic regression, Support Vector Machine (SVM), and Multilayered Neural Network were used for their work. They used these algorithms to predict the restaurant's rating and success and they rounded the predicted ratings to the nearest star. When they used all these algorithms, they observed that among them two algorithms performed much better than others. These two algorithms are Random Forest (60%) and Multilayer Neural Network (56%) algorithm which has about 60% accuracy for binary and 56 % for multi-class classification. Lastly, they also performed sentiment analysis and the accuracy increased up to 85% using several algorithms. For their future Work, they want to find out and add another feature which is types of cuisine to better predict the restaurant's success.

In another research paper [9] by Ibne Farabi Shihab and his fellow teammate’s main focus was to suggest a proper location for setting up a new restaurant business depending on the average rating given by the customer. Here they tried to predict restaurant ratings by using different machine learning algorithms. Their goal was to find a restaurant's rating based on its current feature and then suggest a good location for setting up a new restaurant. For the whole process, the YELP dataset has been used here. They used a linear regression model on the restaurant feature to predict rating. The result was not satisfactory so they used different algorithms like Decision Tree, Logistic Regression, on-Linear SVM for better results. This time the result was much better. After running these algorithms on the restaurant’s features, they observed that Non-Linear SVM highest accuracy score of 97.02% and a precision score of 95.29% which is far better than previous results.

III. DATA COLLECTION AND PREPROCESSING

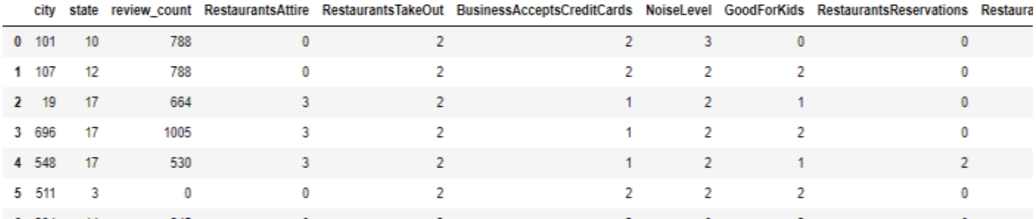
A. *Dataset:*

For any kind of ML system, one has to have a dataset. For our paper, we used the Yelp dataset [11]. As yelp is a globally popular platform for people to rate and review restaurants. The Yelp dataset is huge and has been reviewed by millions of people. Here we are using the freely provided YELP dataset for academic purposes. The main dataset includes six JSON files business, check-in, tips, review, photos, and user.

Another interesting work [10] done by Sunitha Cheriyan and fellow researchers the paper is about Intelligent Sales Prediction Using Machine Learning Techniques .Here data is collected from stores database (2015-2017). Original dataset is consisting of many attributes. These are: Category, City, Type of items and its description, number of items, Quantity, Quarter, Sales, Revenue, Year, SKU description, Week, Year. As the original dataset are so long and that's why non-usable data should be removed from dataset. For predict the sales revenue here they use generalized linear model, Gradient boosted tree, Decision tree. But in the result section GBL performed better than other. The accuracy of this algorithm is 98% within 100% which is much good. As the result is depends on accuracy, precision and recall GBL performed better than other and the value is 50. So, if one wants to performed better than this, he needs to find a strong dataset.

B. *Feature Extraction:*

As we know Yelp data set to have six different sub-datasets from these datasets, we used the YELP academic business dataset for our work here. The business dataset has different columns like business_id, name, address, city, state, postal_code, latitude, longitude, stars, review_count, Is_open, attributes, categories, and hours. From this dataset, we identified and picked only restaurant businesses. There was a total of 63,961 restaurants listed there. As we only picked restaurant business the category column becomes unnecessary so we dropped the column along with some other columns. Name, address, postal_code, latitude, longitude, attributes, and hours were dropped here as they are mainly used to identify the business and not necessary for our work after this, we ended up with 6 different columns. Among these 6 attributes had some nested columns so we separated the attribute column for the dataset and then normalized the attribute column. After normalizing we again found that there are still six nested columns left which are BusinessParking, Ambience, GoodForMeal, DietaryRestriction, Music, BestNights. So, again we separated normalized those columns. We set all NaN variables to string None for data handling purposes later on. There were 33 columns in the attribute after normalizing and extracting the nested columns. And among those six columns that we extracted from attributes BusinessParking had 5 columns, Ambience had 9 columns, GoodForMeal had 6 columns, DietaryRestriction, Music, and BestNights each has 7 columns. So, we ended up with 8 different data frames. Lastly, we merged all the data frames into a single frame and finally ended up with 79 columns. Then we dropped the business_id as there was no further use for it. We then separated stars from the dataset as it serves as our label for this work. Now the remaining columns will act as features for our work. So, finally, we ended up with 63,961 restaurants and 77 features. For rating, we have simplified the rating with 0 as poor, 1 as average, and 2 as good. We also used a label encoder to change the categorical value and assigned them numerical values. Now our dataset is cleaned and ready for applying different algorithms.



	city	state	review_count	RestaurantsAttire	RestaurantsTakeOut	BusinessAcceptsCreditCards	NoiseLevel	GoodForKids	RestaurantsReservations	Restaura
0	101	10	788	0	2	2	3	0	0	
1	107	12	788	0	2	2	2	2	0	
2	19	17	664	3	2	1	2	1	0	
3	696	17	1005	3	2	1	2	2	0	
4	548	17	530	3	2	1	2	1	2	
5	511	3	0	0	2	2	2	2	0	

Fig.1: Dataset after processing

IV. USED ALGORITHMS

We used different machine learning algorithms. For this SKLearn model was used. We split our dataset into 80:20 ratios. Where 80% data were used for training the model and 20% for testing purpose

A. *Decision tree:*

The Decision Tree calculation can use to tackle Regression and Classification issues. It makes a training model which predicts the estimation of target factors by taking in choice guidelines speculating from preparing information. for our work within the tree class of SKLearn [12] we used the decision-tree class. The accuracy we get from the decision tree is 83.6% with a precision score of 86.16%.

B. *Support vector machine (SVM)*

In machine learning, a support vector machine (SVM) is a supervised learning model where the algorithm can give an optimal hyperplane that can categorize new examples when labeled data is given. Svm package from SKLearn was used here and the Accuracy we get from SVM is 91.1%.

C. *k-nearest neighbors' algorithm (KNN)*

K-NN is a Supervised Machine Learning Algorithm. It takes care of both Regression and Classification issues. The accuracy we get from SVM is 91.1% and the precision score is 82.91%.

D. *Stochastic gradient descent (SGD)*

Stochastic Gradient Descent (SGD) is a basic yet exceptionally proficient way to deal with fitting direct classifiers and regressors under raised misfortune capacities, for example, (linear) support vector machines and logistic Relapse. The word 'stochastic' signifies a framework or an interaction that is connected with an arbitrary likelihood. Henceforth, in Stochastic Gradient Descent, a couple of tests are chosen haphazardly rather than the entire informational collection for every cycle. In Angle Plunge, there is a term called "batch" which indicates the all-out number of tests from a dataset that is utilized for computing the slope for every emphasis. Here we get 90.02% accuracy and 86.43% precession.

E. *Gaussian Naive Bayes*

Naive Bayes is a gathering of managed AI classification algorithms dependent on the Bayes hypothesis. It is a straightforward order procedure it has high usefulness. They discover use when the dimensionality of the sources of info is high. Complex characterization issues can likewise be carried out by utilizing the Naive Bayes Classifier. But when the data is continues Gaussian Naïve Bayes perform much better then Naïve Bayes. Here we get 91% accuracy and 82.91% precession.

F. *Convolutional neural network (CNN)*

Although we got relatively good accuracy by using different machine learning algorithms the problem with these algorithms is, they take more time and memory so it was hard to train the full dataset. So, we applied the convolutional neural network here.

In 1980 Yann LeCun first proposed the Convolutional neural network (CNN). It is made out of few layers of artificial neurons. The basic role of these counterfeit neurons is to compute the weighted amount of the sources of info and give an activation value as output. CNN typically comprises a few convolution layers.

CNN takes inputs from a huge dataset and cycles them with irregular qualities. At the point when the output doesn't coordinate with the labels given in the dataset, at that point the model learns and redo the whole process by making corrections in the values. This is basically the whole training process concept for CNN. After a few runs, the models prepare for testing with an unlabeled dataset. At the point when it gets a decent precision on the test dataset, it is prepared for genuine use. In our main CNN model, all the input sequences are made the input of the first embedding layer. Now when the full dataset is embedded, the embedding layer will act as the input layer for the convolutional layer. We used 75 filters in the convocational layer and made the super features which used in the next max-pooling layer. We used global max-pooling for this layer. We applied the SoftMax activation key here. Using this CNN model, we got a 97.22% accuracy score.

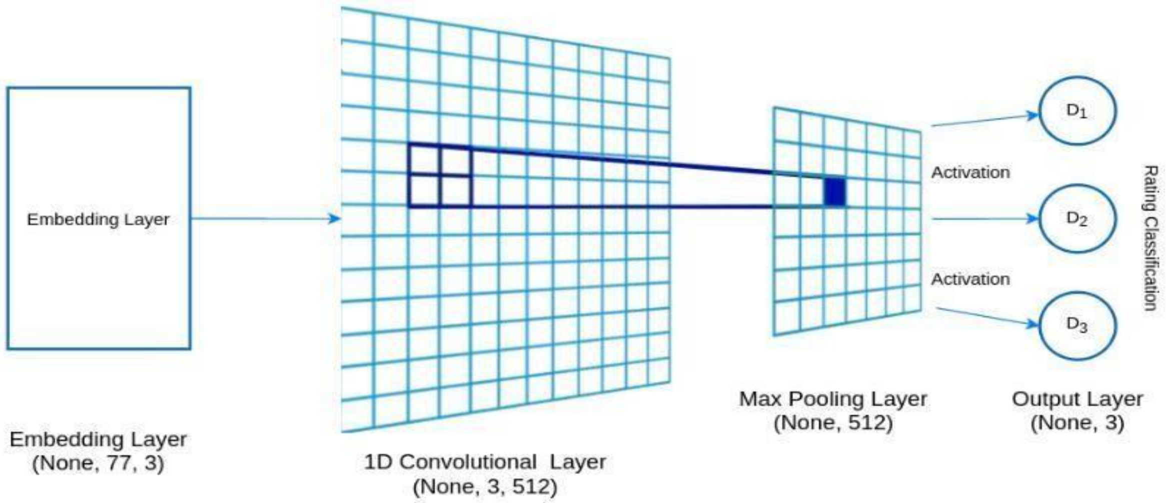


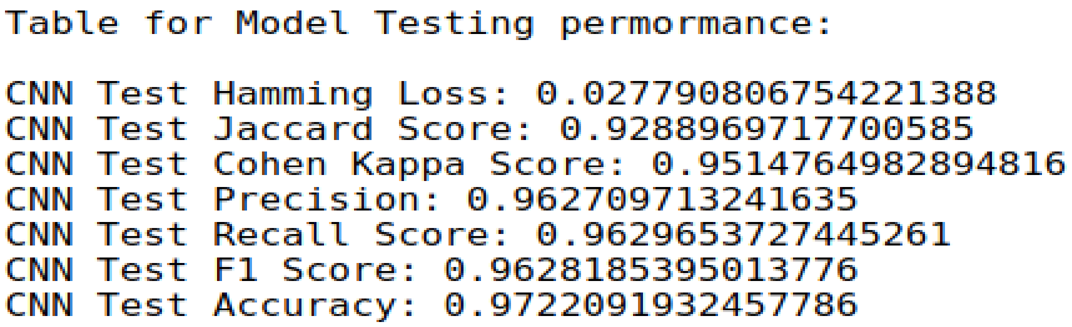
Fig. 2: Convolutional Neural Network (CNN) model.

V. RESULT ANALYSIS

Algorithm Name	Accuracy %	Precision %	Recall %	F1 Score %
Decision Tree	83.6	86.15	83.64	84.81
Support vector machine	91.1	82.91	91.06	86.79
k-nearest neighbors' algorithm	91.1	82.91	91.06	86.79
Stochastic gradient descent	90.02	86.43	90.24	87.62
Gaussian Naive Bayes	91	82.91	91.05	86.79
Convolutional neural network	97.22	96.27	96.3	96.28

Table 1: Different Algorithm performance table

Here, we used multiple machine learning algorithms. Among them, Support vector machine, *k*-nearest neighbors' algorithm, Stochastic gradient descent, stochastic gradient descent, Gaussian Naive Bayes these algorithms are performed average, and the performance of the Decision tree is not so good as others. But when we use CNN, we get a result that is far better than other machine learning algorithms.


```
Table for Model Testing permormance:

CNN Test Hamming Loss: 0.027790806754221388
CNN Test Jaccard Score: 0.9288969717700585
CNN Test Cohen Kappa Score: 0.9514764982894816
CNN Test Precision: 0.962709713241635
CNN Test Recall Score: 0.9629653727445261
CNN Test F1 Score: 0.9628185395013776
CNN Test Accuracy: 0.9722091932457786
```

Table 2: CNN model testing performance

A. *Hamming Loss Jaccard:*

Multi-label classification issues should be assessed victimization different performance measures than single-label classification issues. Two of the foremost common performance metrics square measure acting loss and Jaccard similarity. acting loss is that the average fraction of incorrect labels. Note that acting loss may be a loss operate which the proper score is zero. Jaccard similarity, or the Jaccard index, is that the size of the intersection of the expected labels and therefore the true labels divided by the scale of the union of the expected and true labels. It ranges from zero to one, and one is that the excellent score.

B. *Cohen Kappa:*

Working with unbalanced datasets, Cohen kappa is a valuable estimation metric. While calculating Cohen kappa score, we start with the assumption that the goal and expected class distributions are separate, but that the target class has no bearing on the likelihood of a successful prediction. Cohen proposed that the Kappa outcome be viewed as follows: 0 indicates no compromise,0.01–0.20 indicates zero to mild collaboration, 0.21–0.40 indicates reasonable agreement,0.41–0.60 indicates modest agreement, 0.61–0.80 indicates significant agreement, and 0.81–1.00 indicates almost ideal agreement.

ISSN No:-2456-2165

K= (TA -RA)/(1-RA); TA=(TP+TN)/(TP+TN+ FP+FN) ;

RA=[{(TN+ FP)* (TN+ FN)}+{(TP+ FN)*(TP+ FP)}]/ (TP+TN+ FP+FN)^2

Here, cohen kappa(k),Total Accuracy(TA), Random Accuracy(RA),True Postive(TP), True Negetive(TN), False Positive(FP), False Negetive(FN).

Here, the value of precision, accuracy, recall is much better. Then we use haming loss jaccard and cohen kappa to see how much good our algorithm is. In that case our model perform a great score. The haming loss jaccard value of our algorithm is 0.93 which is very much close to 1 and 1 is the best value for haming loss jaccard. On the other hand the cohen kappa value of our algorithm is 0.95 which is also close to 1. After all, we can easily say that our model CNN perform better then any other previous work in this field. So, we can state that it's thebest model for this type of work.

VI. FUTURE WORK AND CONCLUSION

We use business dataset which is a sub dataset of yelp dataset. Yelp dataset is an USA based dataset. So, our plan is to work with different countries dataset. Here our model is restaurant type business. So, we want to extend it and we will work with different type of business.

Overall, here we tried to make a model which can successfullypredict the expected rating of a restaurant based on its features. Our goal was to predict the rating as accurately as possible. So, after analyzing the performance and results our model is clearly performing much better than any other model. We hope that this model of ours can help a lot of new entrepreneurs in the restaurant business.

REFERENCES

[1.] M. Anderson and J. Magruder, "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Econ. J.*, 2012, doi: 10.1111/j.1468-0297.2012.02512.x.

[2.] G. Parsa, J. T. Self, D. Njite, and T. King, "Why restaurants fail," *Cornell Hotel Restaur. Adm. Q.*, 2005, doi: 10.1177/0010880405275598.

[3.] Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016, doi: 10.1109/CVPR.2016.308.

[4.] S. Khatwani and M. B. Chandak, "Building Personalized and Non Personalized recommendation systems," 2017, doi: 10.1109/ICACDOT.2016.7877661.

[5.] T. Osman, M. Mahjabeen, S. S. Psyche, A. I. Urmi, J. M. S. Ferdous, and R. M. Rahman, "Adaptive food suggestion engine by fuzzy logic," 2016, doi: 10.1109/ICIS.2016.7550755.

[6.] P. A. D. L.Anitha,Kavitha Devi M K, "No Title," *Int. J. Comput. Appl.*, 2013, [Online]. Available: https://www.researchgate.net/publication/260972980_A_Review_on_Recommender_System.

[7.] X. Lu, J. Qu, Y. Jiang, and Y. Zhao, "Should i invest it? predicting future success of yelp restaurants," 2018, doi: 10.1145/3219104.3229287.

[8.] J. Z. A. Wang, W. Zeng, "Predicting New Restaurant Success and Rating with Yelp," 2016.

[9.] F. Shihab, M. M. Oishi, S. Islam, K. Banik, and H. Arif, "A machine learning approach to suggest ideal geographical location for new restaurant establishment," 2019, doi: 10.1109/R10-HTC.2018.8629845.

[10.] S. Cheriyan, S. Ibrahim, S. Mohanan, and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," 2019, doi: 10.1109/iCCECOME.2018.8659115.

[11.] Yelp.com, "Yelp Dataset," 2019. https://www.yelp.com/dataset.

[12.] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12,pp. 2825-2830, 2011.