An Intelligent Approach for Food Recipe Rating Prediction Using Machine Learning

Ismam Hussain Khan
Department of ECE
North South University)
Dhaka, Bangladesh
ishmam.hussain.khan07@gmail.com

Md Habib Ullah Khan
Department of ECE
North South University
Dhaka, Bangladesh
eng.jamee@gmail.com

Md Mamun Howlader
Department of ECE
North South University
Dhaka, Bangladesh
mamunhowlader707@gmail.com

2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) | 978-1-6654-1511-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/CAIDA51941.2021.9425031

Abstract*—In recent times, there are many studies and systems which deal with restaurant rating or individual food rating but rating a recipe using Artificial Intelligence is rare. This study aims to rate recipes based on different attributes using different Machine Learning algorithms. It compares the performance of different classifiers in rating a recipe based on different performance criterion. This can be economically beneficial to restaurants by helping them improve their recipes and getting more customers. It can also be used in a more personal level to improve household recipes and for the customers of restaurants to decide which restaurant is better for a specific dish based on how good their recipe is.

***Index Terms*—Machine Learning, Recipe Rating, Classification, Random Forests, Scikitlearn, Python**

I. INTRODUCTION

Food tasting and rating is becoming more and more popular everywhere which is clearly visible in different online and social media platforms. Food quality can be assessed from different points of view. In this paper we are concerned about rating a food recipe depending on different attributes.

A recipe rating system based on how it is prepared and user reviews can be very helpful for the food industries. Different restaurants can rate their recipes using this and adjust their recipes for more user or customer acceptance. This can also be useful on a personal level for people interested in cooking, experimenting with different recipes and would like to see the rating of their recipes.

There are many factors that play a part in a person's selection of food. Although finding reviews on a particular restaurant, its overall service, food quality is very much available but information about a particular dish or recipe is not as available. If someone is interested in a particular dish or recipe it becomes difficult for them to check how good it is. Different restaurants may also make the same dish using different recipes. So this system can help them see how their dish is rated compared to others.

II. LITERATURE REVIEW

There have been a couple of works based on food rating, restaurant review, calories estimation and food quality rating based on different attributes such as taste, health and more. Here is a list of such papers. [1] is a paper which gives reviews about the Thai restaurants of the world. It follows the approach of extracting review about Thai restaurants from social media and tries to classify them using the help of neural networks

This [2] paper discusses about the main themes of food/restaurant industry including taste, customer's experience, suitable location value from 294,034reviews on Yelp.com using the Latent Dirichlet Allocation(LDA) and also positive view (depending on taste) along with negative view (value). This study also proves a robust classification algorithm based on support vector machine(SVM) and Fuzzy Domain Ontology(FOD) which outperforms other traditional classification algorithms such as Na Naïve Bayes(NB) and SVM in predicting the generosity of online reviews.

[3] is a paper based on applying deep learning on food science. It focuses on applying deep learning as an advanced data mining tool for food sensory research. Their survey indicates that deep learning in food science outperforms conventional machine learning algorithms.

In [4] focuses on rating individual food items of restaurants rather than rating the entire restaurant collecting data from online reviews. They use named entity recognition[NER] technique to

identify individual food names in restaurant reviews.

[5] is a paper which uses a recursive neural network for sentiment analysis task on 'Amazon Fine food review'. They parsed binary trees using Stanford NLP parsers.

In [6] discusses the attributes that have the most effect on online restaurant review and their star rating and finds that food, service and context are the 3 main attributes contributing to the star rating of restaurants.

[7] this paper focuses on classifying fast food restaurant reviews. It quantifies textual reviews based on overall opinion of each online review.

III. METHODOLOGY

A. Data Collection

Data was collected from Epicurious.com using web scraping in python and it was stored in a CSV file.

B. Data Pre-Processing

A python code was applied to remove the unnecessary attributes. After that the exact number of reviews and servings were found by scraping. Total Number of sentences in the instruction column was counted to find the number of instructions per recipe and that number was placed in the instructions column.

After that, we got 5 attributes that are the number of reviews, % Make Again, number of ingredients, number of servings, and number of instructions. The% make again denotes the number of people who would make the dish again after having followed the recipe. After testing the result with each at-tribute and finding their correlation we discarded2 of the attributes and had a final dataset with 3 attributes(Servings,Ingredients, Instructions).

C. Classifier Setup

After discarding unnecessary and irrelevant con-tents in the previous steps, scraping data, dropping missing observations and transforming it into a proper data-set, the total data was divided into 2 sets, keeping 80% data in the training set and 20% data in the test set. After that, different machine learning algorithms i.e Naive Bayes, Logistic Regression, KNearest Neighbour, Decision Tree, Random Forest, Support vector Machine were applied to find training and testing accuracy.

D. Ensemble Learning

Helps to improve machine learning results by combining several models.[8] Here we have combined Decision Tree, Random Forest, and KNN to get the improved result of our prediction.

IV. RESULT AND ANALYSIS

A. Training Accuracy

Out of our 28,954 instances we trained different machine learning algorithms on 23,163 data and discovered that Decision Tree and RandomForest give the best accuracy, precision, recall andF1 scores. After that the most impressive overall performance is yielded by Logistic Regression.

TABLE I
Accuracy Training Results for Classification Models

Classifier	Training Accuracy	Precision	Recall	F1 Score
Naive Bayes	84%	76%	84%	80%
L.Regression	86%	75%	86%	80%
K-NN	86%	80%	86%	81%
Decision Tree	86%	81%	86%	80%
Random Forest	90%	89%	90%	87%
SVM	86%	75%	86%	80%
E.Learning	87%	86%	87%	81%

B. Cross-Validation Accuracy (Training)

In cross validation, SVM, Logistic Regression and KNN yield the highest accuracies and all of them have negligible standard deviation that is +/- 0.00. Naive Bayes has the worst performance compared to any other algorithm and has a high standard deviation as well.

```
NB Accuracy: 0.80966087 (+/- 0.01686057)
LR Accuracy: 0.86607942 (+/- 0.00060466)
KNN Accuracy: 0.86181347 (+/- 0.00380732)
DT Accuracy: 0.82078804 (+/- 0.00747645)
RF Accuracy: 0.84367301 (+/- 0.00678088)
SVM Accuracy: 0.86607942 (+/- 0.00070186)
```

Fig. 1. Accuracy Training Cross-Validation Results for Classification Models

C. Testing Accuracy

We tested our models on 5,791 instances and found that Logistic Regression gives the best training accuracy followed by SVM and KNN. KEN also yields the best precision and F1 scores. So overall, in our training phase KNN works better than other classifiers. However KNN and Logistic Regression yield decent overall performance as well.

TABLE II
Accuracy Testing Results for Classification Models

Classifier	Testing Accuracy	Precision	Recall	F1 Score
Naive Bayes	84%	76%	84%	79%
L.Regression	86%	75%	86%	80%
K-NN	86%	76%	86%	80%
D.Tree	86%	75%	86%	80%
R.Forest	84%	76%	84%	79%
SVM	86%	75%	86%	80%
E.Learning	86%	78%	86%	80%

D. Cross-Validation Accuracy (Testing)

In testing, Logistic Regression, SVM, and KNN performs much better as well with high accuracy and less standard deviation and prove to be the best classifiers for rating recipes.

```
NB Accuracy: 0.80 (+/- 0.03)
LR Accuracy: 0.87 (+/- 0.00)
KNN Accuracy: 0.86 (+/- 0.00)
DT Accuracy: 0.80 (+/- 0.02)
RF Accuracy: 0.83 (+/- 0.01)
SVM Accuracy: 0.87 (+/- 0.00)
```

Fig. 2. Accuracy Testing Cross-Validation Results for Classification Models

V. CONCLUSION AND FUTURE WORK

In this study we have tried to apply different machine learning algorithms to see and compare how well they do in rating a recipe. However, this study isn't only about rating recipes. This Is also an experimental study to see if we can actually rate recipes using machine learning. We had 28,954 instances from there we tested on 20%of them. Since the lowest testing accuracy achieved is 81% after experimenting with different classifiers we are optimistic about that.

Based on what has been covered in this research study we can still make some improvements. More Attributes can be added and analyzed as we have a limited number of attributes. Also with more attributes and data we can try to find out how healthy a recipe is. After collecting more instances we can test how neural networks perform in rating recipes as there have been many studies recently using neural networks on food science.

REFERENCES

[1] Niphat Claypo and Saichon Jaiyen. Opinion mining for thai restaurant reviews using neural networks and mrmr featureselection.In2014 International Computer Science and Engineering Conference (ICSEC), pages 394–397. IEEE,2014

[2] Yi Luo and Xiaowei Xu. Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp. Sustainability, 11(19):5254.

[3] Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu, and YongHe. Application of deep learning in food: A review. Comprehensive Reviews in Food Science and Food Safety,18(6):1793–1811, 2019.

[4]Burusothman Ahiladas, Paraneetharan Saravanaperumal,Sanjith Balachandran, Thamayanthy Sripalan, and SurangikaRanathunga. Ruchi: Rating individual food items in restau-rant reviews.InProceedings of the 12th InternationalConference on Natural Language Processing, pages 209–214, 2015

[5] Jiayu Wu and Tianshu Ji. Deep learning for amazon food review sentiment analysis.

[6] Qiwei Gan, Bo H Ferns, Yang Yu, and Lei Jin. A text mining and multidimensional sentiment analysis of online restaurant reviews.Journal of Quality Assurance in Hospitality & Tourism, 18(4):465–492, 2017

[7] Lindsey Wright. Classifying textual fast food restaurant reviews quantitatively using text mining and supervised machine learning algorithms. 2018

[8] Thomas G Dietterich et al.Ensemble learning. The Handbook of brain theory and neural networks, 2:110–125,2002