International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol.10, No.2, 2023, pp.106-113.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Analyzing Restaurant Reviews to Predict Customer Satisfaction Trends

Dr.N. Baggyalakshmi [a*], V. Anugrahaa [b], Dr.R. Revathi [c]

[a*] *Assistant Professor, Department of Computer Science, PSGR Krishnammal College for Women, India.*
E-mail: baggyanethra@gmail.com
[b] *Student, BSc. Computer Science, PSGR Krishnammal College for Women, India.*
E-mail: grahaaanu20@gmail.com
[c] *Assistant Professor, Department of Computer Science, Karpagam Academy of Higher Education, India.*
E-mail: ravathilakshay@gmail.com

Abstract

Numerous studies comparing the quality of various restaurants' cuisines have been conducted. Area, average cost for two people, votes, cuisines, mostly flavour, and restaurant type are some of the criteria used to evaluate restaurants. Finding out which restaurants people like to eat at and reading reviews of such places is the primary objective here. The goal of this study is to develop a model that can foretell if a review of the eatery will be favourable or negative. A number of prediction algorithms, including Multinomial Naive Bayes, SVC, XGB Regressor, Pipeline, and Logistic Regression, will be utilised to achieve this goal. Finally, we'd like to identify the "best" model that can forecast the reviewer's emotional tone.

Keywords: Naive Bayes, SVC, XGB Regressor, Pipeline, and Logistic Regression, Restaurant.

DOI:10.71086/IAJSE/V10I2/IAJSE1014
Received: 18 March 2023/Revised: 24 April 2023/Accepted: 15 May 2023/Published: 29 June 2023

Introduction

In this Internship Machine Learning with Python using MI it was a cource of making predictions using ml algorithms [1].

Training Program

The internship is a platform where the trainees are assigned with the specific task. In the initial days of the internship, I was trained on the following:

- Python Programming
- Machine Learning Algorithms

Data Set

This section describes, in brief, the data that has been used for the research. Data from restaurant was used in this project, the major amount of data was extracted from public website Kaggle (Kaggle.com), data regarding the review and linked was obtained from a leading Restaurant in India. Data from restaurant sources was integrated together to form a staging data-set [2]. For predicting the review is either positive or negative which uses for the people to say that the which restaurant is best in class and it also uses for restaurant to improve their level k of standarsds in their quality items either it may be the quality food, private space, surrounding of the place, etc.

Dataset Extraction and Transformation

Data related to the Restaurant review was collected in .csv format, the data related to review was extracted using data extraction tool provided by (Mozenda (n.d.)) in .csv files [3]. Data being from public portal had multiple records which got mixing and irrelevant values; data cleaning was performed in Microsoft Excel by collecting all the records to a record having unwanted and missing values. Once the data-set was added to google colab Unwanted columns were left over there, and extracted only wanted (correctly organized) and then divided them into two parts that review and tlinked.1 and then the cleaned data was transformed to be suitable for the model. The original data-set had only the review as a representation of language, to have a consistent metrics for the language score that is either 0 or 1. Similarly, by undertaking the training and testing data we created a prediction model using SVC machine learning algorithm.

Algorithms

Linear Regression

One method in machine learning that relies on supervised learning is linear regression. It runs a regression analysis. Using a set of independent variables, regression models a desired prediction value. Discovering the connection between variables and making predictions are its primary uses. Two main factors that distinguish one regression model from another are the amount of independent variables utilised and the nature of the relationship between the dependent and independent variables.

SVC

The Linear Support Vector Classifier (SVC) method applies a linear kernel function to perform classification and it performs well with a large number of samples. If we compare it with the SVC model, the Linear SVC has additional parameters such as penalty normalization which applies 'L1' or 'L2' and loss function.

Pipeline

- Data transformation and correlation into a model, followed by analysis to provide outputs, can be automated through the use of a pipeline, which streamlines the machine learning workflow [4]. With this ML pipeline, feeding data into the ML model becomes a completely automated operation.
- Separating your machine learning processes into smaller, more manageable pieces that can be reused and re-pipelined to build models is another form of ML pipeline. This ML pipeline eliminates unnecessary steps, making the model-building process faster and easier.
- As a natural outgrowth of the recent trend towards microservices designs, this follows the logic that you can gradually increase the software's power by dividing it into smaller, more specialized components. Linux and Unix, two popular operating systems, were also built on this premise. By

combining simple functions like as 'grep' and 'cat,' powerful pipelined functions can be generated.
In the Internship had three phases:

i. Training Phase
ii. Designing and Development Phase
iii. Testing and Maintenance Phase

As the final task, a main project was developed using machine learning models to predict the chance of a student to be admitted to a master's program. This will assist students to know in advance if they have a chance to get accepted. This project predicts the admission of a student based on different features including university rating, student's undergraduate GPA, GRE score, research experience and etc. This predicts that how much chances are there that the student will get admission in his selected university or not. In this project I have used multiple algorithms including linear regression, artificial neural network (ANN), random forest regressor, decision tree regressor. In the end I have deployed this model on a Web Based GUI to check student's admission chances and these models are working fine

Background Study

Used multiple machine learning models to create a system that would help the restaurant owner to get review that is either positive or negative by predicting using the given review. The secondly it helps the customers to get the best hotel near by his location by seeing the review [5]. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables.

Waters and Miikkulainen (2013) created a review system to help tourists choose the best accommodation. in order for machine learning to model work, we need to input numerical values for these categorical variables. So, to encode Yes/No as 0/1, apply Label Encoding on these two features. Following encoding, divide the dataset in half along the X and Y axes, and then in half again, creating a 70% train set and a 30% test set. Since our features have varying scale ranges, we should standardise the dataset. The values will be brought to a common range after applying standard scaling, making it straightforward for the model to compute and making computation fast. We utilised both logistic regression and support vector machines (SVC) to build the model; both models worked equally well, but we settled on logistic regression for the final system because of how simple it is. Although the admissions committee was able to cut review time for the restaurant by 74%, they still needed human intervention to decide the final status.

Disadvantages of the Background Study

- Limitation of this system only relied on the restaurant makes the restaurant to go down and it takes so much time as if they change their behaviour, quality, surroundings etc.
- The existing system lagged the factor of the research work in the related field.
- This model achieved only 80% accuracy.
- To improve the accuracy, we need to use more number of training data and also weneed to use high performing algorithm

Literature Review

Geler et al [6] the study focuses on sentiment analysis of restaurant service evaluations using machine learning. In order to predict customer satisfaction with restaurants and food services, a comparison of many regression models is given. The Shanghai Lujiazui Commercial Zone food service establishments provided the experimental data, which included keywords taken from the customer reviews. Furthermore, the data have spatial labelling, which makes it possible to do distinct studies for various geographic areas. The meal flavor, service, and environment were the three observed satisfaction factors that could be predicted using the keywords that were taken from the customer reviews.

Adak et al [7] studied explainable artificial intelligence (XAI) techniques, deep learning (DL), and machine learning (ML) models in order to forecast consumer attitudes in the FDS space. A review of the literature showed how popular lexicon-based and machine learning strategies are for predicting customer reviews' sentiments in FDS. Unfortunately, because the decisions taken could not be explained or the model could not be interpreted, there were only a few research using DL approaches. The following are the main conclusions of this systematic review: Although 77% of the models lack interpretability, organizations can make the case for the system's explainability and credibility. While DL models in other areas are accurate, they are not explainable, which can be accomplished by implementing XAI.

Baker et al [8] consumer expectations are expressed in prose in online restaurant reviews, and star ratings reflect total consumer satisfaction. Adversative connectives (ACs) are grammatical constructions (i.e., elements such as but, although, however, and even though) that customers use in language to indicate whether or not their expectations are met. The current study uses a corpus of around 35,000 online reviews for US restaurants to investigate the association between star ratings and consumers' use of ACs. To do this, we combine content analysis, mixed-effects models, and thematic analysis. The findings provide valuable insight into how consumers express their online (dis)satisfaction. To be more precise, statistical modeling shows a significant association between the information that is emphasized by the ACs and the ACs themselves.

Li et al [9] We built a sentiment analysis model for online restaurant reviews by combining Word2vec, Bi-GRU, and the Attention approach. This allowed us to account for context in the reviews and get better results with fewer parameters. As a hidden layer weight update, Bi-GRU ensures symmetry, which is one of its key advantages. In order to train and validate the model, we used restaurant evaluations from Dianping.com. Based on the F1-score, it appears that the Word2vec+Bi-GRU+Attention sentiment analysis model outperforms the commonly used sentiment analysis models by more than 89%. We utilised deep learning techniques to investigate sentiment analysis in online meal ordering platforms with the goal of improving its performance in the restaurant review arena.

Punetha and Jain [10] provides a framework for unsupervised mathematical optimization that can be used to analyze reviews' sentiment and emotions. The suggested model carries out two functions. It does two things: first, it finds the positive and negative sentiment polarity in a review; second, it uses a review to classify customer satisfaction as either satisfactory or unsatisfactory. There are two phases to the framework. Performance scores are initially produced by combining the rating, context, and emotion scores from each review. We implement a non-cooperative game on performance scores in the second stage and reach Nash Equilibrium. The inferred sentiment from the review and the client satisfaction feedback are the outputs of this stage. The trials yielded state-of-the-art findings and were conducted on two restaurant review datasets. Using statistical methods, we verified the results and determined their significance.

Proposed Methodology

The principal objective of the research is to help the restourants to get there level of standards in a way that positive or negative reviews and also helps to who are aspiring to go visit the restaurant. The Restaurant review Prediction system will help them to evaluate the chances of success in improving the customers needs. It will help them in saving a huge amount of time and money spent in the knowing each and every customers decisions. Also, it will help them to limit the number of customers liking the restaurant and what the customers are expecting from the restaurant and it also helps the customers by suggesting them the best Restaurants where they have high chances of their needs.

Advantages of the current study

- Information about the prediction analysis is clear to enter all the required information to predict the review is either positive or negative.
- The user interface code will interact with the Linear Regression, KNN, SVC to provide the users with the required result.
- Lower-quality restaurants may either close their doors or raise their standards in reaction to shifts in customer demand, since user ratings may drive customers to more upscale establishments.

Data Flow Diagram

The machine learning models are trained with the given dataset. The machine learning models used in this project are linear regression, linear simple vector classifier(svc), random forest regressor, decision tree regressor. Once the models are trained, the model are entered to predict the chances of getting positive or negative review.

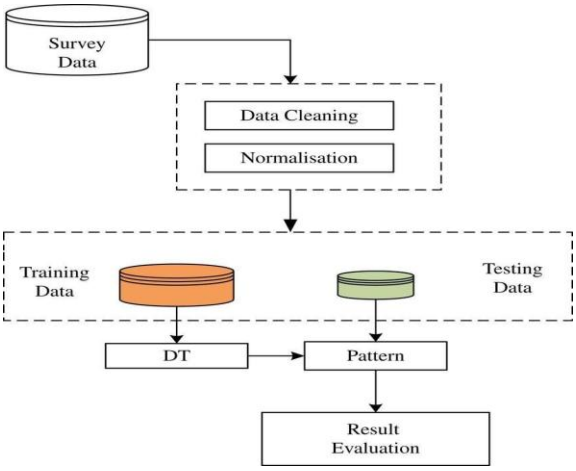


Fig. 1: Dataflow Diagram of Restaurant Review Prediction

Modules in the System

Modules

1. Exploratory Data Analysis in Machine Learning
2. Data Visualization
3. Training and Testing
4. Train and Evaluate Linear Support Vector Classifier
5. Train and Evaluate pipeline in machine learning

Modules Description

Exploratory Data Analysis: Exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. EDA is used to understand data, find patterns, detect outliers, and test hypotheses with the help of summary statistics and graphical representations. EDA is often the first step in data analysis and can be used to identify relationships between variables, uncover errors in data, and guide the selection of appropriate statistical models.

Data Visualization: To help the human mind digest and grasp the provided data more easily, I used data visualisation to summarise the data using graphs, photos, and maps. While data visualisation is beneficial for representing small and large data sets, it really shines when dealing with massive datasets when human eyes simply can't handle the sheer volume of information.

Training and Testing: Datasets are divided into two parts for this project. The training data is the initial subset; it is a subset of our real dataset that is used to train the machine learning model and find patterns. Our model is trained in this manner. The second batch is called the testing data.

Train and Evaluate Linear Support Vector Classifier (SVC): A linear kernel function is used to do classification in the Linear Support Vector Classifier (SVC) approach, which works effectively with a high number of data. If we compare it with the SVC model, the Linear SVC contains additional parameters such as penalty normalisation which applies 'L1' or 'L2' and loss function.

Train and Evaluate Pipeline in Machine Learning



Fig. 2: Machine Learning Pipelines

Software Details

System Requirements and Specification

Table 1: Hardware Requirements

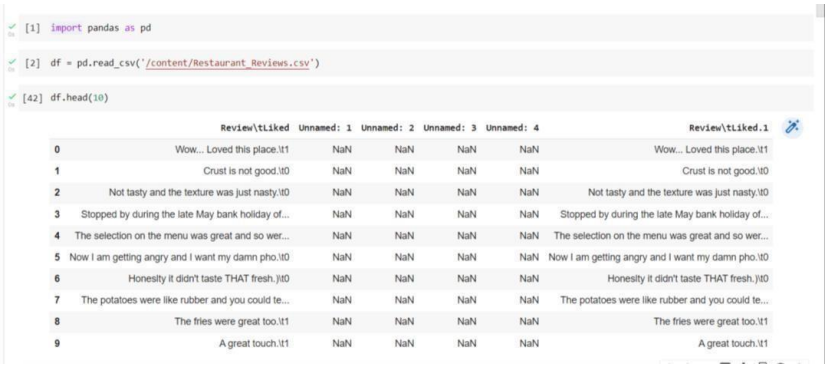
Processor	x86 or x64
Hard disk	216 GB or more
Ram	512 MB (minimum), 1GB (recommended)

Table 2: Software Requirements

Operating System	windows or Linux
Development Environment	Anaconda Navigator (Jupiter Notebook or Spyder)

Results and Discussion

Screenshots



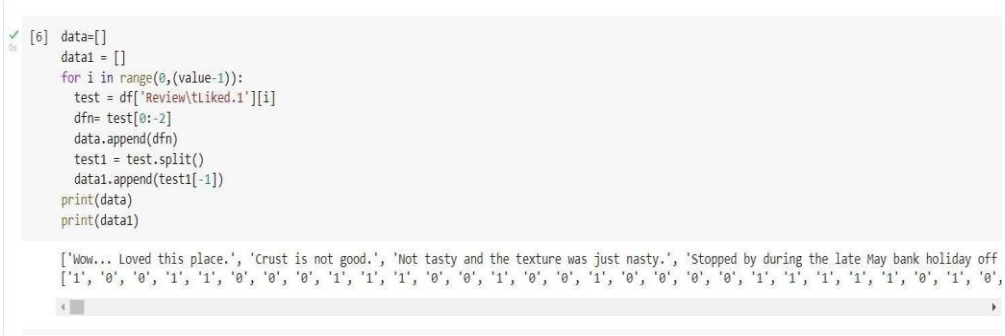
```
[1] import pandas as pd

[2] df = pd.read_csv('/content/Restaurant_Reviews.csv')

[42] df.head(10)
```

	Review\tLiked	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Review\tLiked.1
0	Wow... Loved this place.\t1	NaN	NaN	NaN	NaN	Wow... Loved this place.\t1
1	Crust is not good.\t0	NaN	NaN	NaN	NaN	Crust is not good.\t0
2	Not tasty and the texture was just nasty.\t0	NaN	NaN	NaN	NaN	Not tasty and the texture was just nasty.\t0
3	Stopped by during the late May bank holiday of...	NaN	NaN	NaN	NaN	Stopped by during the late May bank holiday of...
4	The selection on the menu was great and so wer...	NaN	NaN	NaN	NaN	The selection on the menu was great and so wer...
5	Now I am getting angry and I want my damn pho.\t0	NaN	NaN	NaN	NaN	Now I am getting angry and I want my damn pho.\t0
6	Honesity it didn't taste THAT fresh.)\t0	NaN	NaN	NaN	NaN	Honesity it didn't taste THAT fresh.)\t0
7	The potatoes were like rubber and you could te...	NaN	NaN	NaN	NaN	The potatoes were like rubber and you could te...
8	The fries were great too.\t1	NaN	NaN	NaN	NaN	The fries were great too.\t1
9	A great touch.\t1	NaN	NaN	NaN	NaN	A great touch.\t1

Fig. 3: Importing Data



```
[6] data=[]
    data1 = []
    for i in range(0,(value-1)):
      test = df['Review\tLiked.1'][i]
      dfn= test[0:-2]
      data.append(dfn)
      test1 = test.split()
      data1.append(test1[-1])
    print(data)
    print(data1)

['Wow... Loved this place.', 'Crust is not good.', 'Not tasty and the texture was just nasty.', 'Stopped by during the late May bank holiday off
['1', '0', '0', '1', '1', '0', '0', '0', '1', '1', '1', '0', '0', '1', '0', '0', '1', '0', '0', '0', '0', '1', '1', '1', '1', '1', '0', '1', '0',
```

Fig. 4: Cleaning Data



```
[14] from sklearn.model_selection import train_test_split
     x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.05 ,random_state=0)

     print(x_train,y_train,x_test,y_test)

'1' '1' '1' '0' '0' '1' '1' '1' '0' '0' '1' '0' '1' '1' '1' '0' '0' '1'
'0' '1' '0' '0' '1' '0' '0' '1' '0' '0' '0' '0' '1' '0' '0' '1' '0' '1'
'0' '1' '1' '0' '1' '0' '1' '1' '1' '0' '0' '0' '0' '0' '0' '1' '0' '1'
'0' '1' '1' '1' '0' '1' '1' '1' '0' '0' '1' '0' '1' '1' '0' '0' '0' '0'
'1' '0' '0' '0' '1' '1' '0' '1' '1' '1' '0' '0' '0' '0' '1' '0' '1' '1'
'1' '1' '0' '0' '1' '0' '1' '0' '1' '0' '1' '1' '0' '1' '1' '0' '0' '1'
'1' '0' '0' '0' '1' '0' '1' '0' '1' '1' '1' '1' '0' '0' '0' '0' '0' '0'
```

Fig. 5: Dividing Data Set into Training and Testing Data

```
[17] from sklearn.feature_extraction.text import CountVectorizer

     vect = CountVectorizer(stop_words='english')
     x_train_vect = vect.fit_transform(x_train)
     x_test_vect  = vect.transform(x_test)

[18] from sklearn.svm import SVC
     model = SVC()
     model.fit(x_train_vect,y_train)

     SVC()

[19]
     y_pred = model.predict(x_test_vect)
     y_pred

     array(['1', '0', '0', '1', '0', '1', '0', '1', '1', '0', '0', '0', '1',
            '0', '0', '0', '0', '0', '1', '0', '0', '1', '1', '0', '1', '1',
            '0', '1', '0', '1', '1', '0', '1', '0', '0', '0', '0', '1', '1',
            '0', '0', '0', '0', '1', '1', '1', '1', '0', '0', '1'],
           dtype=object)
```

Fig. 6: Fitting the Model and Predicting the Output

```
from sklearn.metrics import accuracy_score,classification_report
accuracy_score(y_pred,y_test)
```

```
0.8
```

Fig. 7: Accuracy of the Data

```
[23] #for a highly imbalanced dataset, not only consider accuracy_score

     print(classification_report(y_pred,y_test))
```

	precision	recall	f1-score	support
0	0.95	0.68	0.79	28
1	0.70	0.95	0.81	22
accuracy			0.80	50
macro avg	0.82	0.82	0.80	50
weighted avg	0.84	0.80	0.80	50

Fig. 8: Report

```
plt.bar(x_test,y_pred)
plt.xlabel("Actual review")
plt.ylabel("Predicted output")
plt.title("Actual review vs Preicted output")
plt.show()
```

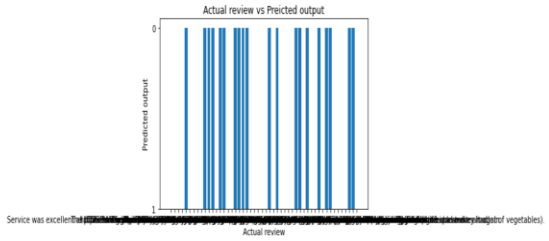


Fig. 9: Output in Bar Graph

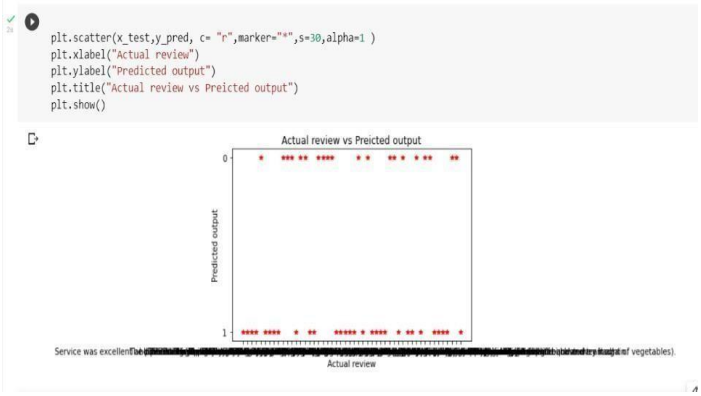


Fig. 10: Output in Scatter Plot

Reference

Pleerux, N., & Nardkulpat, A. (2023). Sentiment analysis of restaurant customer satisfaction during COVID-19 pandemic in Pattaya, Thailand. *Heliyon*, *9*(11), 1-15.

Luo, Y., & Xu, X. (2021). Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic. *International Journal of Hospitality Management, 94,* 102849.

Zibarzani, M., Abumalloh, R. A., Nilashi, M., Samad, S., Alghamdi, O. A., Nayer, F. K., ... & Akib, N. A. M. (2022). Customer satisfaction with Restaurants Service Quality during COVID-19 outbreak: A two-stage methodology. *Technology in Society, 70*, 101977.

Chua, B. L., Karim, S., Lee, S., & Han, H. (2020). Customer restaurant choice: an empirical analysis of restaurant types and eating-out occasions. *International journal of environmental research and public health, 17(17),* 6276.

Polat, S.A. (2022). Analysis of Fine Dining Restaurant Reviews for Perception of Customers' Restaurant Service Quality. *Journal of Tourism & Gastronomy Studies*, *10*(1), 11-28.

Geler, Z., Savić, M., Bratić, B., Kurbalija, V., Ivanović, M., & Dai, W. (2021). Sentiment prediction based on analysis of customers assessments in food serving businesses. *Connection Science*, *33*(3), 674-692.

Adak, A., Pradhan, B., & Shukla, N. (2022). Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. *Foods, 11(10),* 1500.

Baker, M. J., & Hashimoto, B. (2024). Expression of customer (dis) satisfaction in online restaurant reviews: The relationship between adversative connective constructions and star ratings. *International Journal of Business Communication, 61(1),* 148-180.

Li, L., Yang, L., & Zeng, Y. (2021). Improving sentiment classification of restaurant reviews with attention-based bi-GRU neural network. *Symmetry, 13(8),* 1517.

Punetha, N., & Jain, G. (2023). *Game theory and MCDM-based unsupervised sentiment analysis of restaurant reviews. Applied Intelligence (Dordrecht, Netherlands*), 1.