iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** XII **Month of publication:** December 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.48109

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue XII Dec 2022- Available at www.ijraset.com*

# Restaurant Review Analysis using Natural Language Processing

Pankaj Kunekar[1], Haripriya Arya[2], Abhishek Dighekar[3], Atharv Bagade[4], Harish Garud[5], Eissa Abdelbari[6]
[1, 2, 3, 4, 5, 6]*Vishwakarma Institute of Technology Pune, India*

***Abstract:*** ***The most effective tool any restaurant can have is the capability to track the daily sales of their food and beverage. Currently, recommendation systems plays an important role in both academia and industry. These are very helpful to manage an overload of information. In this paper, applied machine learning techniques for user reviews were used and valuable information in the reviews were analyzed. For both the customers and the owners, reviews are useful to make data-driven decisions. We built a machine learning model with Natural Language Processing techniques which captures a user's opinions from user's reviews. A lot of businesses fail due to the lack of profit and a lack of proper improvement measures. Mostly, restaurant owners face a lot of difficulties to enhance their productivity.***

***Keywords: Restaurant reviews, Python, Machine Learning, NLTK, Pandas, Naive-Bayes Algorithm, NLP, Survey***

## I. INTRODUCTION

A restaurant's profits are based upon how many customers do they serve daily. There has to be an average number of customers who come to the restaurant on a daily basis. If we were to really know what they felt about the restaurant, it would be of great help for the restaurant owner to know about their customer-base and understand what kind of changes are to be made and how those changes can be made. A restaurant itself can conduct surveys where they might be able to get feedback from their customers and know what might be lacking or what can be changed in order to make a greater profit and retain the same customers while bringing in new ones too.

It would mean a huge task for a restaurant which might be serving more than a 100+ customers daily. The restaurant owners can look at the surveyed data and make changes according to them, but it might take a lot of time to read the surveys, understand the problems, make decisions and then to concoct a strategy best suits them. This process can certainly work for smaller restaurants.

In this project, the first step would be to collect the data from the customers. This step can be done by taking an online survey where the data is then stored on a cloud platform. This data can be pre-processed which involves elimination of stop words from the reviews such as "the", "and", "is" etc. Then the stemming method will be applied which will remove all the affixes and prefixes which would help the model to understand the base word clearly. The vectorization of the text information will be performed. Count vectorizer and TFIDF vectorizer will then be used. The machine learning algorithm will then be used to find the base combination of vectorizer method and ML model.

## II. LITERATURE REVIEW

In an approach which was used, they split the pre-processing stage into 3 major steps. The first step was to remove the punctuation marks in the sentences. The special characters like forceful mark and quotes area unit removed by creating an applicable regular expression. The result of this step would be lines that contain only alphabetical characters. The second step was to remove the stop-words from reviews. Stop-words are words that don't generate any feeling or sentiment but are used as connectors in the English language. Words like of, with, the, and are also excluded while using natural language processing (NLP) techniques like Lexical analysis, linguistics analysis, syntactic analysis, revealing integration and pragmatic analysis are applied on the data set to identify and eliminate stop-words. Linguistics analysis removes the stop-words like "not". In opinion mining, the presence or absence of the word "not" plays an important role. For example, if a review says "the crust isn't good", the removal of stop-words will result this sentence into "crust good". Therefore, a negative opinion can change into a positive one. To keep away from such drawbacks, they changed the linguistics analysis step and created a certain method that does not remove such stop-words. The third step in pre-processing is to convert the initial words to their root words. Root words are words without a prefix or a suffix. For example, love is the root for words like loves, loved, affectionate, etc. As they were interested only in original opinion or sentiment, such conversion made the task easier. The Porter Stemmer formula was applied to change all the words in the data set into their root words.

Machine Learning isn't a brand new technique for text process. numerous researchers applied machine learning techniques for eating place reviews classification.

©IJRASET: All Rights are Reserved 

A hybrid classification model for sentiment analysis of restaurant reviews was used. They used a classifier that contains a support vector machine and Naive Bayes models. With their model, they achieved an accuracy of ninetieth. Sasikala.P [2] projected a model for classifying restaurant reviews victimization sentiments in the words. Their model relies on the score combined with existing text analyzing packages. the majority use 'yelp' for locating a decent restaurant. Yelp reviews ar terribly helpful for locating a decent restaurant. Boya Yu[3] projected support vector machines for analyzing eating house options victimization Sentiment Analysis on Yelp Reviews. Kirange [4] conjointly projected a Support Vector classifier for feeling Classification of Restaurant Reviews. They compared their model with Naive Bayes, K-NN and neural network models and shown that SVM achieved smart results. Tri Doan [5] et.al proposed a variant of on-line random forest classifiers for activity sentiment analysis on user reviews. They showed that their model achieved an accuracy almost like offline methods. Ekaterina Pronoza[6] et.al projected a eating house info extraction technique for the eating house recommendation system. Veda Waikul [7] et.al projected an SVM classifier for classifying restaurant reviews. With the model that they developed, they achieved an accuracy of 77%.

## III. PROPOSED METHODOLOGY

The proposed system here can distinguish the review data-set into positive and negative reviews. To develop this system the incorporated packages used are Pandas, NLTK, TKinter.

### *A. NLTK*

The Natural Language Toolkit (NLTK) is used to build programs in Python that work with data on human language that applies statistical natural language processing (NLP). It holds text processing libraries that tokenize, parse, classify, stem, tag and also performs semantic reasoning. As the data set contains both, positive and negative reviews, a classifier was built which is able to classify a new given review into positive or negative. Classification is a supervised machine learning problem. It specifies the classes where the data elements belong and is optimally used when the output has discrete and finite values.

### *B. Pandas*

Pandas is a software library, which is written for Python which can perform data manipulation and analysis. It provides users with data structures and operations to manipulate numerical tables and time series. We imported the restaurant review data set using the Pandas library

### *C. TKinter*

Tkinter is used in Python to create Graphical User Interfaces (GUI's). It is a framework built in the Python standard library. The GUI for this project was built through Tkinter.

### *D. Vectorization*

In NLP, words or phrases from vocabulary to a corresponding vector of real numbers can be mapped, which can be used to find word predictions, word similarities/semantics. Converting words into numbers is a process called vectorization. To remove special characters and to break a setence down and converting all the characters into lowercase count vectorization was used. Then we used the TF-IDF method, which helped us measure the importance of a given word based on the frequency of their appearance in a document.

### *E. Process Description*

We built a classifier that classifies a given a review into positive and negative. A data set was imported then the removal of stop-words and lemmitization took place. This would help us in a massive way as removal of the stop words can help the model learn the classification faster and can be more effective and efficient. To find word predictions, count vectorization was used where the phrases from reviews correlate with the vectors. These vectors were then fed into the model which was then ‘trained’. Now predictions using test data can also be generated. The accuracy score and classification report can thus be generated. Now if words like “bad”, “worst” and “stale” were to be present in the training data-set then the tested data will decide the review to be negative as it has been trained to do so. When the model gives out the result to be “1” then it is supposed to be a positive review and if it were to be “0” then it is supposed to be a negative review.

Multiple classification algorithms were used such as Logistic Regression, Bernoulli Naive Bayes, Multinomial Naive Bayes. We got the best results with Multinomial Naive Bayes algorithm.

## IV. RESULTS

Successfully able to install all libraries required for the project. Gathered around 1000 datasets having both positive/negative reviews. Figure 1. shows the model accuracy of the project having precision, recall, f1-score and accuracy. Figure 2. shows the accuracy level with a live data set that was classified 3. describes the rating on average of the restaurant's review analysis.

| Algorithm | The ratio of training versus testing dataset | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 80:20 | | | | 75:25 | | | |
| | Sentiment | P% | R% | F% | Acc.% | P% | R% | F% | Acc.% |
| Naïve Bayes | Positive | 92 | 91 | 92 | 91 | 92 | 93 | 92 | 92 |
| | Negative | 91 | 92 | 92 | | 92 | 92 | 92 | |

Figure [1]: Model Accuracy

The system is effective at detecting positive and negative reviews that are partially written or are not very sensible. In order to distinguish between a positive and negative review it takes into account the words which are most used to reduce its decision making time for the reviewing process.

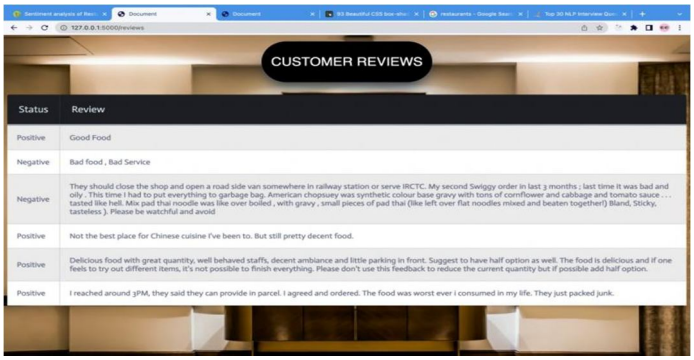



Figure [2]: Output 1

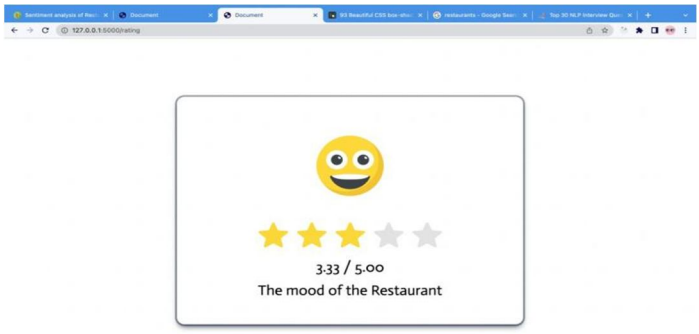



Figure [3]: Output 2

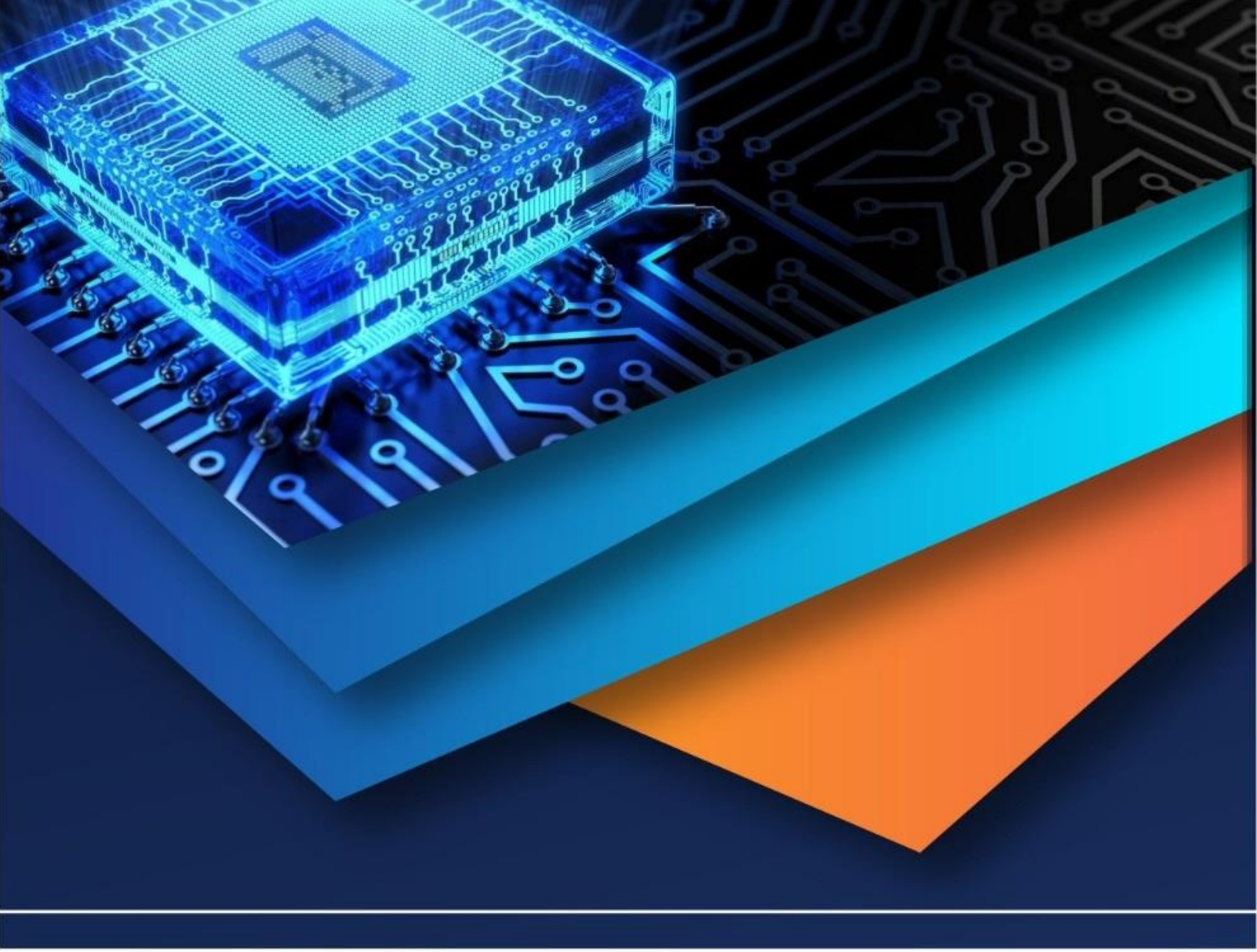
Wrong letters in a word which would not make sense in the real world and a lack of clarity are the main problems the system encounters. However, analyzing the given data will always draw two conclusions, either good or bad.

## V. CONCLUSION

Learning about the customer is the first and foremost priority for any type of business to be successful and have a happy customer. The same is the case with restaurants. Restaurants who need to know their customer-base can greatly benefit from this project. We proposed machine learning NLP techniques for classification of restaurant reviews . We apply stemming for efficiency. The motive of the model is to detect the sentiments/opinions of the reviews. The developed model has an accuracy of 77.67% and it successfully identifies the sentiments of the reviews. The model was tested with a data set that contained reviews from customers and it detected the sentiments accurately. We can conclude that the intention of this project was fulfilled and the model can be used to identify the sentiments of the reviews.

## REFERENCES

[1] M. Govindarajan, "Sentiment Analysis of Restaurant Reviews Using Hybrid Classification Method", International Journal of Soft Computing and Artificial Intelligence, ISSN: 2321-404X, Vo-2, Iss-1, May-2014
[2] P. Sasikala, L.M. I Sheela,"Sentiment Analysis of Online Food Reviews using Customer Ratings", International Journal of Pure and Applied Mathematics, Volume 119 No. 15, 3509-3514,2018.
[3] Boya Yu, Jiaxu Zhou, Yi Zhang, Yunong Cao, "Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews", arxiv,20 Sep 2017
[4] D.K.Kirange and 2Dr. Ratnadeep R. Deshmukh, "Aspect and Emotion Classification of Restaurant and Laptop Reviews Using Svm", International Journal of Current Research Volume. 8, Ise-03, pp. 28352-28356, March, 2016.
[5] Tri Doan, Jugal Kalita, ''Sentiment Analysis of Restaurant Reviews on Yelp with Incremental Learning", 2016 15th IEEE International Conference on Machine Learning and Applications.
[6] Ekaterina Pronoza, Elena Yagunova, Svetlana Volskaya, and Andrey Lyashin, "Restaurant Information Extraction (Including Opinion Mining Elements) for the Recommendation System", MICAI 2014, Part I, LNAI 8856, pp. 201–220, 2014, Springer International Publishing Switzerland 2014.
[7] Veda Waikul, Aruna Pavate, Onkar Ravgan, "Restaurant Review Analysis and Classification Using SVM," IOSR Journal of Engineering (IOSR JEN), ISSN (e): 2250- 3021, ISSN (p): 2278-8719, PP 49-52.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)