iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36540

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: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Restaurant Review Analysis using NLP

Tanya Bhatt[1], Sarthak Sharma[2], Tushar Gaur[3], Shubham Aggarwal[4], Dr. Sweeta Bansal[5]
[1, 2, 3, 4]*Student, Dept. of Computer Science, Inderprastha Engineering College, Uttar Pradesh, India*
[5]*Professor, Dept. of Computer Science, Inderprastha Engineering College, Uttar Pradesh, India*

Abstract: ***Because starting a business, such as a restaurant, is not easy. Reading all of the evaluations and comparing input to other restaurants becomes challenging and time consuming for a restaurant owner. As a result of this analysis, it is an important factor to consider when making a dining selection. This paper provides an efficient restaurant review prediction model to predict the review for the restaurant based on a certain collection of client reviews. Restaurant reviews are labelled as 0 or 1, with 0 denoting a negative review and 1 denoting a positive one. The findings of this study could be employed as a long-term marketing strategy for review website creators, allowing customers to sort and filter useful reviews based on their preferences.***

Keywords: Review prediction, Sentiment Analysis, Support Vector Machines(SVM), Natural Language Processing(NLP)

I. INTRODUCTION

When visiting a new restaurant, customers want to know what to expect. Who better to inform them than a prior client? The more people who hear about your restaurant, the more likely they are to visit. In this day of globalisation, we are constantly on the lookout for new ideas that will save us time, simplify tasks, and eliminate the need for manual processing.

Customers at restaurants rate and leave reviews based on their level of satisfaction. Other customers can use these ratings and reviews to assist them decide whether or not to visit those eateries. These ratings can also help restaurant owners make changes based on customer feedback in order to improve their business. Textual information can be found in restaurant reviews. Machine learning is one of the subset of artificial intelligence (AI). ML allows systems to learn without having to be explicitly designed, and this learning can be used to solve issues. Machine learning takes data as input and learns essential relationships from it in order to make judgments that meet the needs of the user. The learning process begins with observations such as samples and direct experience, and then looks for patterns in the data to help make better judgements about how to forecast or categorise new things in the future. Natural language processing (NLP) skills are provided by machine learning for text processing. Using NLP approaches, we can simply analyse our textual datasets. Data analysts can use NLP to apply machine learning and deep learning algorithms to their textual datasets. For classifying reviews, we employ machine learning algorithms.

II. RELATED WORK

There is a discussion on classification performance. There are several articles on related issues, such as recommendation systems (Adomaviius, G and et al., 2005), informative peer-prediction approach (Nolan Miller and et al., 2005), and rating prediction (Adomaviius, G and et al., 2005). G. Adomaviius and colleagues (2005) provide an overview of preferred systems. It also outlines the current version of recommendation systems, which are classified into three categories: content-based, collaborative, and hybrid approaches. These approaches, however, have limits. This paper covers a number of potential improvements that can improve recommendation capabilities and make recommendation systems applicable to a wider variety of applications. Michael J and et al., (2007) offer us with a basic content-based recommendation system that suggests an item based on its description and the user's interest profile. The ultimate recommendation is determined by the combination of these two elements. Although the specifics of an item may differ throughout recommendation systems, there are certain commonalities. For instance, the ability to compare item characteristics.A more similar example can be found in Gayatree Ganu and et al., (2009). Computers have a hard time analysing, understanding, and aggregating free-text reviews. This research introduces new ad-hoc and regression-based recommendation algorithms that take into account the textual component of user reviews in order to identify the information in the text reviews.There are three types of sentiment classification approaches that have been utilised previously. Machine learning algorithms, link analysis methods, and score-based methodologies are examples of these. Pang et al., 2002 conducted groundbreaking research on the efficiency of machine learning approaches when used to sentiment classification tasks.Zhang, Ziqiong, et al., (2011) employed naive machine learning approaches. Bayes and SVM are used to automatically identify user reviews as good or negative in the realm of online Cantonese-written restaurant reviews. The consequences of different feature displays and sizes.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

III. LITERATURE SURVEY

Customer happiness is a critical subject in the fields of marketing and consumer behaviour research. When hotel customers receive outstanding service, they will spread the word to their friends and family. Understanding internet reviews may also help business owners and marketers obtain a better understanding of their customers' attitudes and behaviour, which can be leveraged to improve service and establish a lasting competitive advantage. Prior research has looked into internet reviews to better assess consumer experience and satisfaction in a variety of service contexts, such as the hotel business, short-term rental industry, The airline sector, as well as the wellness industry. According to reports, 87 percent of clients will not consider firms with bad ratings, and 92 percent will rely on internet reviews to judge whether or not a business is good. Sentiment analysis is a technique for determining the author's feelings on a certain subject. A review's sentiment analysis is an examination of a product's opinion. The foundation of sentiment analysis is the use of Natural Language Processing (NLP), text analysis, and certain computational components to extract or delete unneeded elements in order to evaluate whether a statement is negative or positive.A number of online platforms have offered mechanisms for other users to evaluate online reviews . Customers may vote on reviews that they find useful in their decision-making process on sites like Amazon.com and TripAdvisor. The amount of helpful votes may indicate the message's quality. As a differentiating technique, retail website developers might provide useful reviews to enhance website visitors. Researchers used to manually do text analysis to determine the most significant product or service characteristics to customers based on word frequency a few decades ago. Computer learning for textual data analysis, which allows a machine to extract and categorise online reviews, has been used to give additional insights and generate predictions from large quantities of reviews in order to better understand elements that contribute to an useful review.Machine learning approaches for text data are less time consuming and labour expensive than traditional types of manual content analysis. Additional information such as semantics, structures, sequences, and context around neighbouring words is also provided. Text mining classification is a fundamental text-mining job that classifies unstructured input with appropriate categories from a predetermined collection. Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KMN) and SVM with PCA are some of the most often utilised machinery learning approaches for classification and regression analysis. Although NB provides less accurate prediction outputs, researchers prefer it because of its fast processing speed on huge data. SVM is one of the most successful approaches for categorising unlabeled data at the moment. As a result, comparisons between different machine learning algorithms should be conducted in order to identify which data-mining method in the restaurant sector gives the most precision and accuracy.

IV. METHODOLOGY

The Dataset is a.tsv file containing 1000 reviews in text format obtained from www.kaggle.com. The SVM classifier is trained using this dataset. It is divided into two parts: 70% training data and 30% test data. There are two columns in the dataset. The first column comprises written reviews from various people that are connected to the restaurant's food as well as an overall assessment of the establishment. The mood, or whether the review is good or negative, is displayed in the second column. The number 1 denotes a good review, while the number 0 denotes a bad review. Import the dataset and transform it to a Pandas Dataframe. The model should be able to tell whether the review will be good or negative. The Data Frame is then provided as input to the Countvectorizer for further processing after the dataset is cleaned from 1000 reviews and the reviews that are not suitable are deleted from the dataset.

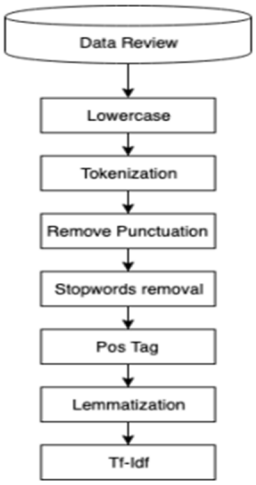
A. Support Vector Machine(SVM) Algorithm

The research has presented a Machine Learning model that would aid in the categorization of restaurant evaluations. To categorise the reviews, this model was trained using the SVM (Support Vector Machine) Algorithm.SVM stands for Supervised Learning and is used to solve both classification and regression issues.However, it is mostly utilised in Machine Learning for Classification issues. SVMs can efficiently do non-linear classification as well as linear classification utilising a technique or parameter called Kernel, which implicitly translates their inputs into high-dimensional feature spaces. Soon, we'll learn more about the Kernel.Unsupervised Learning is another term for SVM. When data is unlabeled, supervised learning is impossible, thus an unsupervised learning technique is necessary, in which data is clustered naturally into groups and new data is mapped to these groups.The support-vector clustering method was developed by Hava Siegelmann and Vladimir Vapnik to categorise unlabeled data using support vector statistics produced in the support vector machines algorithm. It is one of the most commonly used clustering algorithms in industrial applications.

©IJRASET: All Rights are Reserved

B. Natural Language Processing(NLP)

NLP is a branch of machine learning that involves a computer's capacity to comprehend, interpret, modify, and perhaps produce human language. Natural Language Processing (NLP) is a technique that allows computers to comprehend human language. Virtual assistants like Google Assist, Siri, and Alexa are probably the most well-known instances of NLP in action.NLP deciphers and translates human language into numbers, such as "Hey Siri, where is the nearest restaurant?" to make it easier for robots to interpret. Chatbots are another well-known use of NLP, since they may assist you in solving problems while conducting natural language generation — in other words, conversing in plain English.The field of natural language processing (NLP) is concerned with how computers comprehend and translate human language. Machines can understand written or spoken material and execute tasks such as translation, keyword extraction, topic categorization, and more using natural language processing (NLP).



C. QT Designer

Qt Designer is a Qt tool for creating and building Qt Widget-based graphical user interfaces (GUIs). You can use a what-you-see-is-what-you-get (WYSIWYG) approach to create and personalise your windows or dialogues, and test them in a variety of styles and resolutions. Using Qt's signals and slots technology, widgets and forms produced with Qt Designer interact seamlessly with programmed code, allowing you to simply assign behaviour to graphical objects. All Qt Designer properties can be updated dynamically in the code. You may also utilise your own components with Qt Designer thanks to features like widget promotion and custom plugins.Qt Designer is a GUI building tool included with the PyQt installation. A GUI interface may be quickly developed without needing to write code using its simple drag and drop interface. It is not, however, an IDE like Visual Studio. As a result, Qt Designer is unable to debug and build the application.

D. PyQt5 Framework

PyQt5 is the most recent version of Riverbank Computing's GUI widgets toolkit. It's a Python interface for Qt, a powerful and widely used cross-platform GUI library. PyQt5 is a programming language that combines Python and the Qt library. PyQt API is a collection of modules that includes numerous classes and functions. The QtGui module contains all the graphical controls, while the QtCore module contains non-GUI functionality such as interacting with files and directories. Also there are modules for working with XML (QtXml), SVG (QtSvg), and SQL (QtSql), among other things.

E. ZOMATO API

RESTful API from Zomato. It provides you with the most up-to-date and comprehensive restaurant content to use in your apps. It includes 1.5 million restaurants in 10,000 locations around the world. This RESTful API finds restaurants by name, cuisine, or location, and provides detailed information such as ratings, location, and cuisine. By getting the access key, you can gain free and immediate access to restaurant information.The Zomato Web-API utilizes standard HTTP communication and RESTful ideas are used. The HTTP body of requests and responses contains JSON. UTF8 is always used for character encoding. Ascertain that your client can handle SSL/TLS HTTP requests.

©IJRASET: All Rights are Reserved

F. Zomato API Query

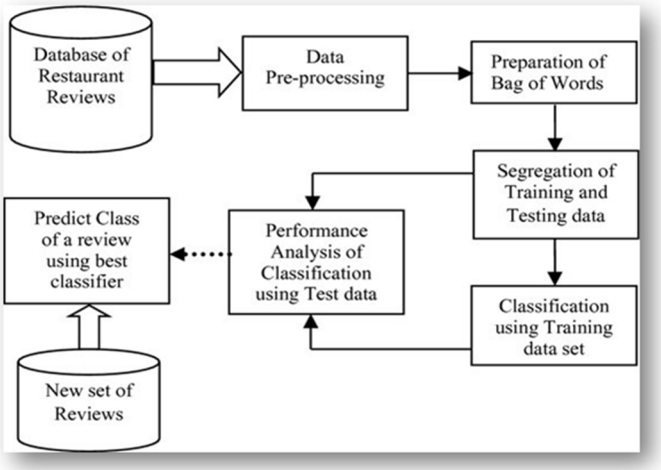
```
1. get city_id using -> /cities API
**********************************************************************************************************************************************
-       query:

curl -X GET --header "Accept: application/json" --header "user-key: a9a841e35c194c4ed583ccd2c4932f22" "https://developers.zomato.com/api/v2.1/cities?q=kanpur"

-       Response:

        {
          "location_suggestions": [
                {
                  "id": 23,
                  "name": "Kanpur",
                  "country_id": 1,
                  "country_name": "India",
                  "country_flag_url": "https://b.zmtcdn.com/images/countries/flags/country_1.png",
                  "should_experiment_with": 0,
                  "has_go_out_tab": 0,
                  "discovery_enabled": 0,
                  "has_new_ad_format": 1,
                  "is_state": 0,
                  "state_id": 0,
                  "state_name": "",
                  "state_code": ""
                }
          ],
          "status": "success",
          "has_more": 0,
          "has_total": 0,
          "user_has_addresses": true
        }
----------------------------------------------------------------------
        here id denotes city_id.
```

G. Proposed System



1) Data from a Restaurant Review Database Pre-processing
2) Preparation of the Bag of Words, then data separation for training and testing.
3) Classification with the use of a training data set
4) Classification Performance Analysis Using Test Data
5) SVM classifier is used to predict reviews.

V. IMPLEMENTATION AND TESTING

This section provides information about the execution condition and highlights the significant advancements made in the use of the dataset to show signs of increased precision in predicting restaurant ratings.

A. Hardware requirements

The following hardware was used for the implementation of the system:
1) 4 GB RAM
2) 10GB HDD
3) Intel 1.66 GHz Processor Pentium 4

B. Software requirements

The following software was used for the implementation of the system:
1) Windows 10
2) Python 3.8.0
3) MacOS9 x orAbove

©IJRASET: All Rights are Reserved

C. Implementation Steps

In this part, we'll go through the approaches that were used throughout the procedure. We'll present some proof for the data we used to make our predictions. We based our decision on data on the website.

1) Restaurant Reviews.tsv is a dataset from Kaggle datasets that contains 1000 restaurant reviews.
2) The website data is saved in a CSV file, from which the data is extracted for prediction.
3) The gathered data is then preprocessed, and a bag of words modelled for the SVM Classifier is created.
4) Next, training and test data are split in a 7:3 ratio; training data will be used to train the classifier that will predict the best value; test data will be used to train the classifier that will predict the best value; considered for the prediction.
5) The Classifier will now be put to the test using the test data, and the accuracy will be reported.
6) The app uses Zomato's API to get real-time reviews, which are then fed into the SVM classifier for prediction, which gives us information about that specific restaurant.

VI. RESULTS

The dataset is used to train the SVM model, and the positive and negative review outcomes are saved in the database.
The greatest prediction accuracy was obtained by combining Natural Language Processing (NLP) with the SVM classification method (77 %)

.

Steps to obtain the accuracy of the model:

1) Initially, we import all the packages which can be implemented in our project.

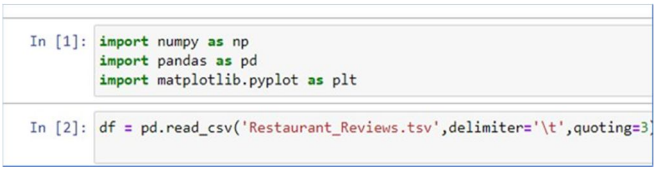


Fig -1: Importing the required packages

2) The data sets will be loaded for testing and training

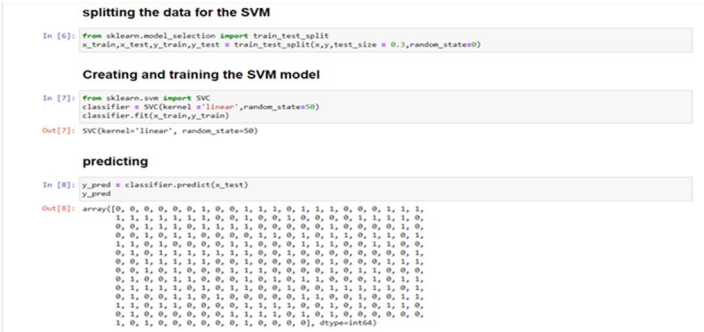


Fig -2: Loading the data set

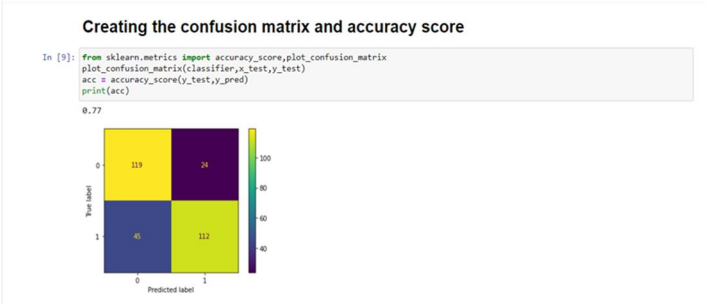
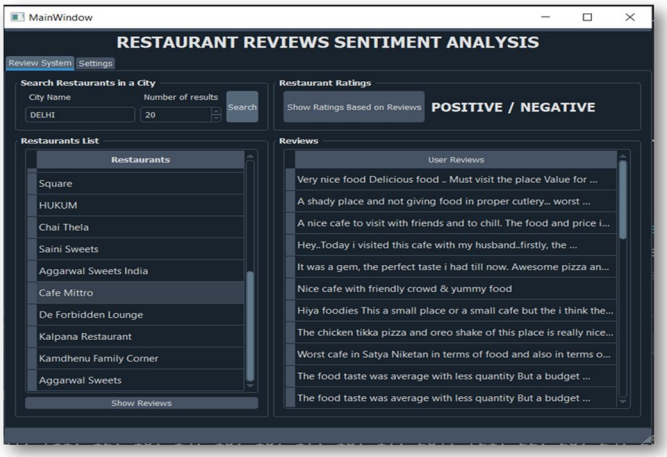


Fig -3: Accuracy of the model

©IJRASET: All Rights are Reserved

3) The accuracy of the SVM classification method attained the best prediction accuracy, according to our findings (77%)

4) Following the creation of the machine learning model, live reviews from Zomato are gathered and input into the SVM Classifier, as seen in the front end window. The classifier will forecast how many good reviews there are and how many negative reviews there are. Finally, it will provide us with the overall rating of the chosen restaurant.



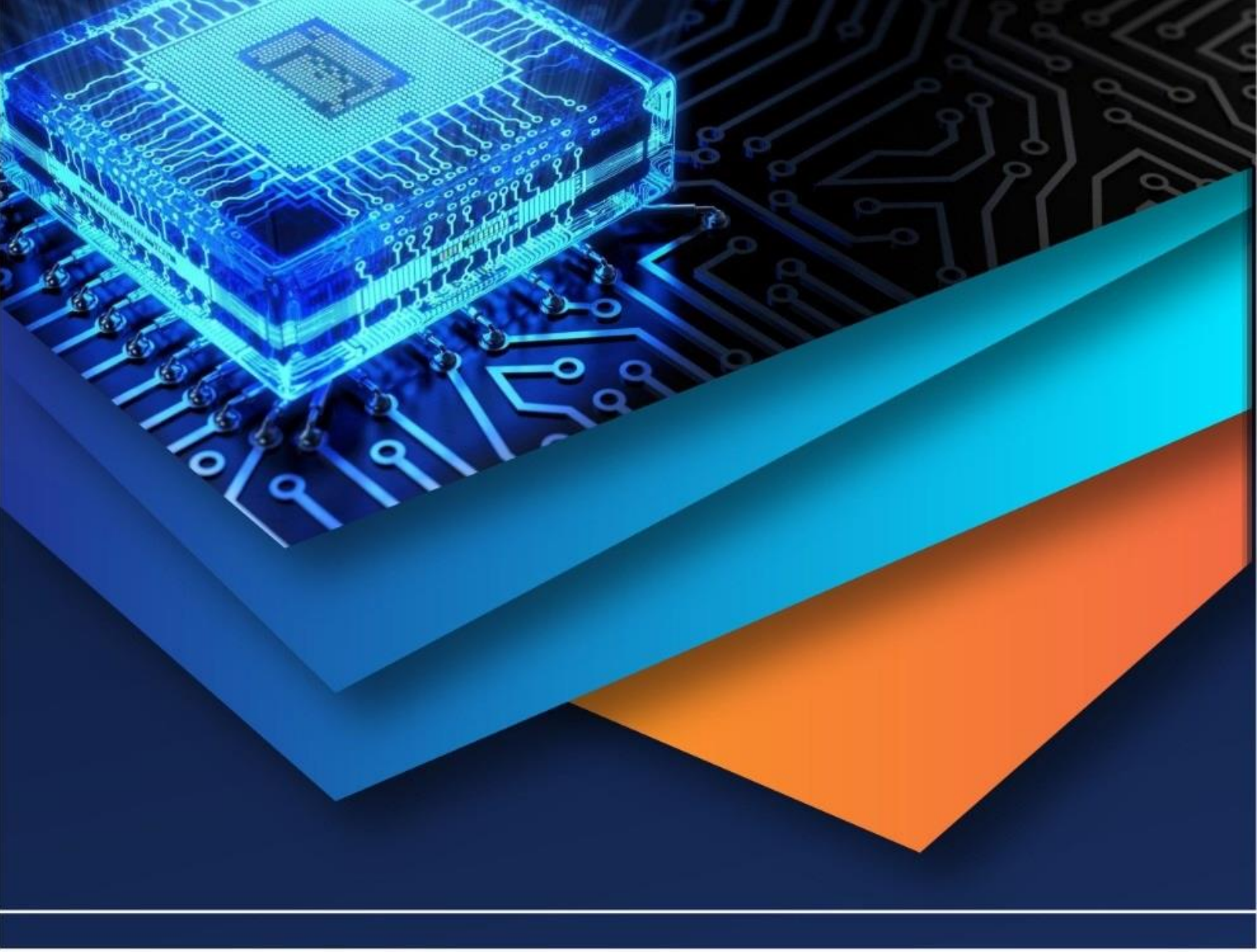
VII. CONCLUSION

This study presents an efficient restaurant review prediction algorithm to predict the review for the restaurant based on 1000 customer reviews. Restaurant reviews are denoted by the numbers 0 and 1, with 0 denoting a negative review and 1 denoting a positive one. Natural Language Processing (NLP) combined with the SVM classification method yielded the greatest prediction accuracy-(77%). This strategy will aid business owners in anticipating consumer feedback and improving the customer experience. As a result of this study, it is an important factor to consider when making a dining selection. It's an important component of the planning process when starting a business like a restaurant.

REFERENCES

[1] Basant, A., M. Namita, B. Pooja, and Sonal Garg. "Sentiment analysis using common-sense and context information." (2015).
[2] Tri Doan, Jugal Kalita, ‘‘Sentiment Analysis of Restaurant Reviews on Yelp with Incremental Learning”, 2016 15th IEEE International Conference on Machine Learning and Applications.
[3] 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.
[4] Priyantina, Reza Amalia, and Riyanarto Sarno. "Sentiment Analysis of Hotel Reviews "International Journal of Intelligent Engineering and Systems 12, no. 42019): 142-155
[5] 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.
[6] 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.
[7] Yan Zhu, Melody Moh, Teng-Sheng Moh, ―Multi-Layer Text Classification with Voting for Consumer Reviews‖, 2016 IEEE International Conference on Big Data (Big Data), Pg 1991-1999
[8] Sanjukta Saha, Dr.A. K. Santra , ―Restaurant Rating Based on Textual Feedback, 2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)
[9] S. Prakash, A. Nazick, R. Panchendrarajan, M. Brunthavan, S. Ranathunga, A. Pemasiri, ―Categorizing Food Names in RestaurantReviews‖, 2016 Moratuwa Engineering Research Conference (MERCon)
[10] Zhenhai Mu, Zhongxuan Tan, Ge Zhu, ―Android based order recommender system‖, 2018 International Conference on Robotics & Intelligent System, Pg 258-261
[11] M Takbir Hossain, Md. Ismail Hossain, Ms. Samia Nawshin, ―Machine Learning Based Class Level Prediction of RestaurantReviews‖, 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh, Pg 420– 423
[12] Toshihiko Yamasaki, Masafumi Yamamoto, and Kiyoharu Aizawa, ―Review-Based Service Profiling and Recommendation‖, 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS), Pg 377-380
[13] Aggarwal, Charu C. "Text Summarization." In Machine Learning for Text, pp. 361-380. Springer, Cham, (2018)
[14] Ocher, Richard, et al. "Semi-supervised recursive autoencoders for predicting sentiment distributions." Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, 2011.

©IJRASET: All Rights are Reserved



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)