ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 1, pp: (7-15), Month: January - March 2024, Available at: www.researchpublish.com

PREDICTION MODEL, SENTIMENT ANALYSIS AND RECOMMENDATION SYSTEM ON ZOMATO BANGALORE DATASET WITH XAI TOOLS LIME AND SHAP

[1]Mekhala Vinod Purohit, [2]Dr. Shantha Rangaswamy

Dept of Computer Science and Engineering,

RV College of Engineerin

Bangalore, India

DOI: https://doi.org/10.5281/zenodo.10725858

Published Date: 29-February-2024

Abstract: **The study of prediction models, sentiment analysis, and recommendation systems has gained prominence in business and application research. It is essential to streamline the suggestion process based on customer needs as internet reviews of restaurants gain traction. This study analyses the emotional tone of online reviews using sentiment analysis and recommendation systems, and provides restaurants with insightful information about how well their products are received by customers. The key features of the study include deriving the factors that led to the model prediction, sentimental tone analysis using the explainable AI tool LIME, and SHAP. This will provide detailed insight into the key features that define the model and the driving forces behind its output.**

Keywords: **prediction; sentiment analysis; recommendation; LIME; SHAP.**

I. INTRODUCTION

The customer community's reliance on product reviews and ratings while ordering and shopping online is significant. Every choice, from placing an order on an e-commerce website to selecting a mouthwatering dish on Zomato, is dependent on ratings and reviews

Considering this, the most noteworthy finding in this study is a proposed method by which the restaurant's owners and Zomato data analysts can comprehend the relationship between the actual ratings given by customers and the predicted values generated by the employed algorithms, which is further supported by the **LIME** explainer tool.

The tone of the reviews is examined for **sentimental analysis** in addition to the **prediction model**, and the XAI tool **SHAP** is added to support the analysis performed by the model. The confusion matrix for balance and imbalance data is calculated for the sentiment analysis model.

This model also contains a **recommendation engine** to take into account the entirety of the review's world. The consumer can input the approximate cost for two people and the cuisine they would like to dine, and a recommendation is made based on the inputs provided.

II. RELATED STUDY

The referred research papers collectively delve into the domains of recommender systems, sentiment analysis, and data-driven insights in the context of restaurant reviews and user preferences. Even though each article offers distinctive insights, similar aspects including dataset application, methodology, and potential future use become evident.

Utilization of Datasets:

Several publications utilize sizable datasets gathered from a variety of sources, including Yelp, Zomato, Kaggle, and user-generated reviews, to support their analysis. Understanding user preferences and restaurant evaluations is based on these datasets.

Techniques:

A wide range of techniques, including machine learning algorithms like Naive Bayes, Support Vector Machines, Random Forest, and more, are used across the publications. To glean useful insights from text data, methods including sentiment analysis, content-based filtering, collaborative filtering, and deep learning architectures like CNN-LSTM are used. A fuller knowledge of user opinions is also made possible by methods like Latent Dirichlet Allocation (LDA) and aspect-based sentiment analysis.

Outcomes from existing models:

The accuracy rates of these researches show a wide range, with varying findings. When it comes to sentiment classification and recommendation tasks, some articles achieve great accuracy, while others investigate the possibilities of hybrid models. Notably, one article demonstrates the potential for accurate user sentiment interpretation by achieving a 97% accuracy rate in sentiment analysis. Additionally, user-specific recommendations are delivered via collaborative filtering and recommendation algorithms, proving their usefulness.

The future possibilities from these studies:

In their individual studies, the scholars have a strong emphasis on future developments and advancements. Extending datasets, investigating new features, and adapting models for different languages are all aspects of future efforts. Improvements are sought by using cutting-edge methods like rule-based sentiment analysis and adding user ratings and reviews as features. Additionally, recommendations for group dining scenarios, ingredient traceability using blockchain, and further hybrid strategies are being explored to enhance the precision of recommendations. Researchers hope to keep improving their models in order to increase client retention and satisfaction.

These research pieces demonstrate the enormous potential of data-driven techniques, machine learning, and sentiment analysis in improving restaurant suggestion and review interpretation processes, optimizing user experiences, and increasing satisfaction.

However, we cannot understand the reasoning behind the behavior of the model. XAI tools can aid us with it. In our model, we included the XAI tools for the same.

III. PROBLEM STATEMENT

Zomato, having a big data set, makes exploratory data research on it challenging. We are developing a model for prediction, sentiment analysis, and recommendation engine, to the benefit of both restaurant/company owners and end consumers. The proposed techniques here are Naive Bayes Text Classifier, Support Vector Machine Algorithm, and Explainable AI LIME tools

IV. OBJECTIVES

i. To carry out prediction on ratings, sentiment analysis on the ratings and reviews submitted by the users.

ii. To create a recommendation system that allows consumers to choose a restaurant that suits their preferences.

iii. To utilize the LIME and SHAP tool to deduce all of the parameters influencing the model's predictions for better analysis of the reviews on the results.

V. SCOPE

Zomato is an online resource that offers details about cafes, bars, and restaurants along with user reviews and ratings. Zomato's ratings prediction, sentiment analysis and recommendation engine can aid users and restaurant owners by offering insightful information. Here are a few potential application of the model.

1. ***Sentiment Analysis***

The ability to analyse the tone of user evaluations for each restaurant using sentiment analysis can be adapted to the website/official app. This can assist users in determining the restaurant's advantages and disadvantages and assist them in choosing where to dine. In order to spot patterns in user evaluations and pinpoint areas where restaurants need to improve, Zomato can also employ sentiment analysis not just on the customer side of the operations but also for commission gauging with restaurants.

2. ***Personalized Recommendations***

Zomato Zomato's recommendation engine enables it to offer consumers tailored recommendations based on their dining tastes and history. Zomato can recommend eateries that are likely to appeal to each user by examining user data. Although the current model includes this feature, it doesn't justify why the selection was made. That is where our model comes into the picture, which uses XAI tools to deduce the analysis behind the prediction.

VI. BASIC FLOW

3. *Model training and testing:* This stage divides the dataset into training and testing sets based on the columns "Reviews_List" and "Rate" from the Zomato dataset. The accuracy rate for the same is computed using the Naive Bayes Text Classifier Algorithm. This stage generates prediction ratings for restaurants, which is used to compare the actual rating to the forecast rating.

4. *Prediction Model:* The prediction of restaurant ratings using Naive Bayes text classification is a valuable task that leverages machine learning techniques to assess and categorize user reviews into different rating classes. Naive Bayes is a probabilistic algorithm that, in this case, analyzes the textual content of user reviews to predict the corresponding restaurant rating, typically on a scale such as 1 to 5 stars.

5. *Analysis Model:* Support Vector Machines (SVM) is a potent machine learning approach that may be used for sentiment analysis of the Zomato dataset to identify and classify the sentiments expressed in user evaluations. SVM, a powerful classification algorithm, excels at taking on sentiment analysis jobs. Since user-generated restaurant reviews are most likely included in this dataset, SVM is used to categorise them into 'positive,' 'negative,' or 'neutral' sentiment categories.

Data pre-processing, which includes text cleaning, tokenization, and feature extraction, usually starts the process. Utilising the method TF-IDF (Term Frequency-Inverse Document Frequency) vectorization, text reviews are frequently converted into numerical representations. Then, an SVM classifier is trained using these numerical features.

6. *LIME Tool*: LIME operates by changing the feature values of these instances and tracking changes in the predictions made by the model. LIME develops explanations by comparing the model's predictions with perturbed data to its predictions with the original data.

LIME offers justifications in the form of feature weights. It reveals which characteristics had the most impact on the rating predictions made by the model for the chosen instances. It might show, for instance, that elements such as restaurant ambiance, food style, or location had the most influence on the anticipated ratings.

7. *SHAP Tool:* The SHAP (SHapley Additive exPlanations) tool can be a valuable resource for gaining insights into the factors that contribute to sentiment analysis of Zomato ratings. When applied to sentiment analysis of Zomato ratings, SHAP can help answer questions like, "What features or words in user reviews have the most significant influence on sentiment predictions?" and "How do different features contribute to the overall sentiment?”

To determine SHAP values for each instance in the dataset, we use the SHAP tool. The contribution of each feature to the model's prediction for a particular instance is shown by the SHAP values.

Examine the SHAP values to comprehend the significance and influence of many features in establishing sentiment. Visualisations like summary graphs or descriptions of each individual case might aid in understanding the findings.

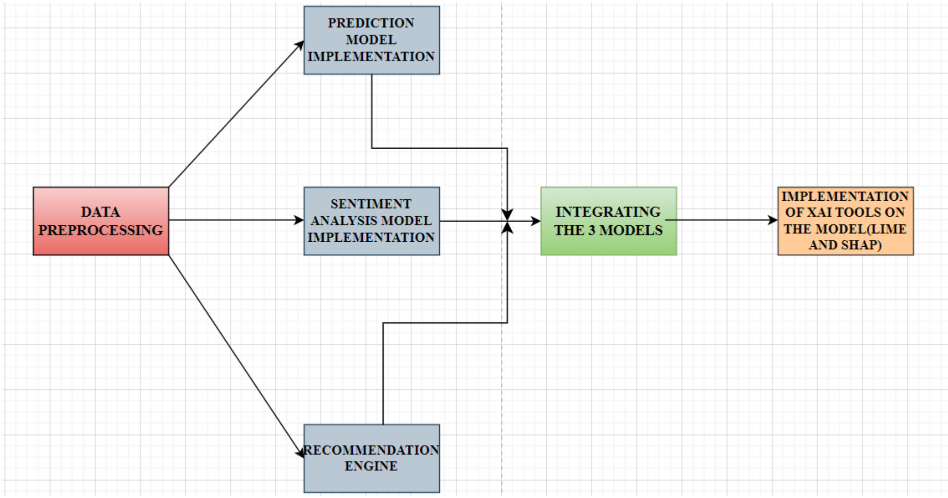

Fig vi. A flow of the hybrid model.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

1.Prediction model

This predicts the ratings for each restaurant as well as the elements that influence the prediction. Users of restaurants can use this to analyse the ratings, allowing them to make better decisions. The predicted and actual values are stored in a new CSV file called "Zomato_test_results.csv". Some snippets of the outputs obtained are shown below.

The validation accuracy and testing accuracy for prediction model is about 39% using Naïve Bayes text classification algorithm.

1.1 The graph predicts the actual and predicted ratings, showcasing the relativity between them for the first few instances of the dataset.

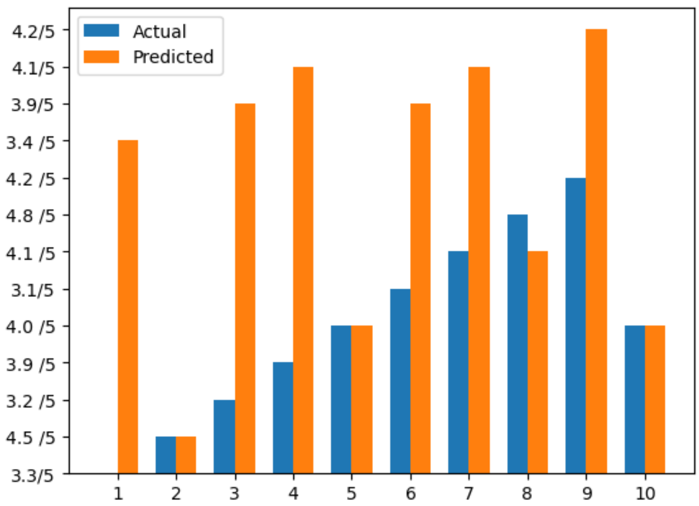

Fig 1.1: A plot of actual v/s predicted ratings

1.2 The prediction model's accuracy for validation is 38.84% and testing is 38.95% using Naïve Bayes Text Classification Algorithm.

Validation Accuracy: 38.84457857296832

Testing Accuracy: 38.95236042250485

1.3 Few instances from the result file of prediction. A csv file is generated with instance number, the predicted value and the actual value.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 1, pp: (7-15), Month: January - March 2024, Available at: www.researchpublish.com

PREDICTED RATING	ACTUAL RATING
3.4 /5	3.3/5
4.5 /5	4.5 /5
3.9 /5	3.2 /5
4.3 /5	3.9 /5
3.9/5	4.0 /5

Fig 1.3: A table of actual v/s predicted ratings

1.4 LIME (Local Interpretable Model-agnostic Explanations) explainer tool is used to analyze feature contributions for predicting ratings in the Zomato dataset. Within a loop that iterates over test instances, LIME is employed to generate explanations for the model's predictions. These explanations reveal how specific features influence the rating predictions. After predicting the ratings, the code displays interactive LIME explanations for each instance, highlighting the significant feature contributions. Subsequently, it extracts and visualizes these contributions using horizontal bar charts, where feature names are plotted against their respective contributions. This approach enables a clear and interpretable representation of the factors that most heavily impact the model's rating predictions, aiding in understanding the model's decision-making process in the context of the Zomato dataset.

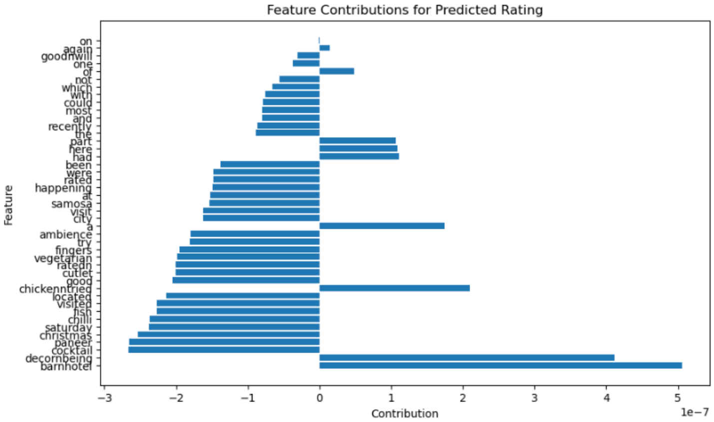

Fig 1.4: Feature contribution chart using LIME explainer

2. <u>Sentiment analysis model</u>

2.1 A bar graph is plotted which depicts number of instances under each of the sentiments (Positive, Neutral, Negative) for Zomato Bangalore dataset instances.

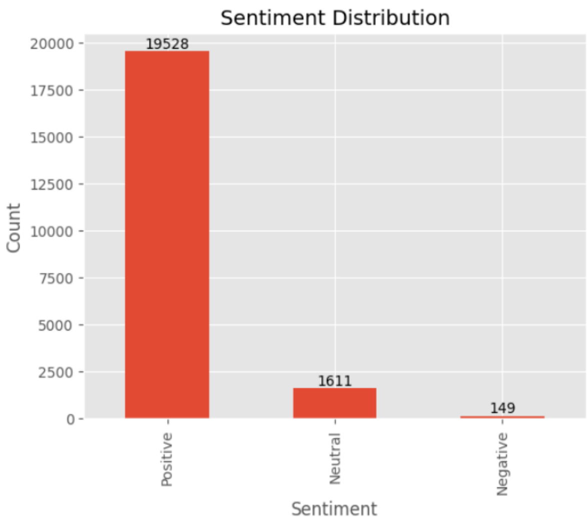

Fig 2.1 A bar graph depicting number of instances under each of the sentiments (Positive, Neutral, Negative) for the dataset

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 1, pp: (7-15), Month: January - March 2024, Available at: www.researchpublish.com

2.2 The confusion matrix for validation and testing of sentiment analysis is given in the table below. The imbalanced data is being considered as well.

TESTING SET REPORT (IN %)				
	PRECISION	RECALL	F1-SCORE	SUPPORT
NEGATIVE	91	77	83	13
NEUTRAL	98	66	79	152
POSITIVE	97	100	99	1964
ACCURACY			97	2129
MACRO AVG	95	81	87	2129
WEIGHTED AVG	97	97	97	2129

VALIDATION SET REPORT (IN %)				
	PRECISION	RECALL	F1-SCORE	SUPPORT
NEGATIVE	100	58	74	12
NEUTRAL	93	76	84	161
POSITIVE	98	100	99	1956
ACCURACY			98	2129
MACRO AVG	97	78	85	2129
WEIGHTED AVG	97	98	97	2129

Fig 2.2 Precision, Recall, F1-score table

2.3 SHAP (Shapley additive explanations) tool is used to generate SHAP values, which help explain the model's predictions. It aims to provide insights into how different features contribute to sentiment predictions for various classes (Negative, Neutral, and Positive). For each sentiment class, it creates a separate SHAP explainer tailored to that class and calculates SHAP values for the filtered validation data. These SHAP values represent the impact of different features on sentiment predictions for that specific sentiment category.

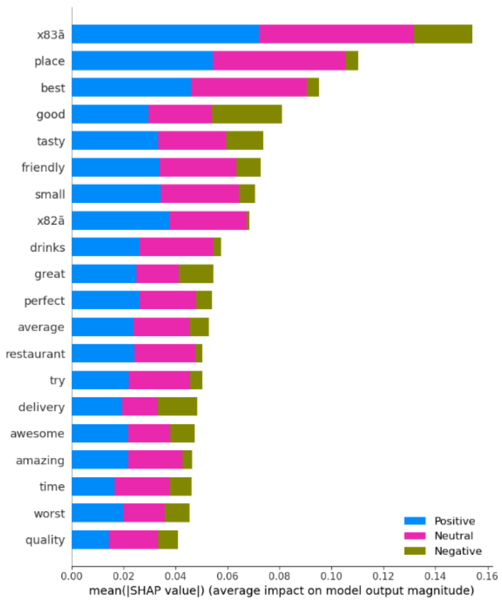

Fig 2.3 SHAP tool output for one of the reviews

3. Recommendation model

The input taken by the recommendation system is the type of cuisine, and the cost for two peoplethat the user wishes to spend. Based on this it recommends a restaurant that is in favour of user's input.

3.1 The below images show few examples of the same

```
Enter your cuisine preference: Mexican
Enter your budget for two people: 700
Recommended Restaurant: Stajasia Ahara
```

```
Enter your cuisine preference: Cafe
Enter your budget for two people: 1500
Recommended Restaurant: Savera Cafe
```

Fig 3.1 User input based recommendation

3.2 A graph is plotted to understand the approximate cost distribution for two people. The frequency of the distributed is charted out to understand the relationship between the two.

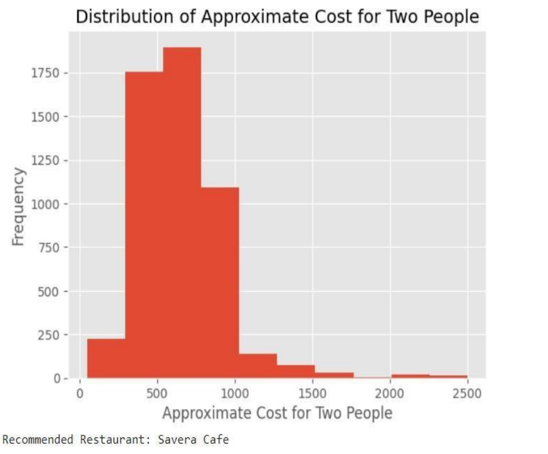

Fig 3.2 Plot of approximate cost in Recommendation System

3.3 The main purpose of this code is to create a pie chart to visualize the top 10 cuisine preferences among Zomato users based on the data in the dataset. It achieves this by first counting the occurrences of each cuisine type in the 'cuisines' column of the dataset and selecting the top 10 cuisines with the highest counts.

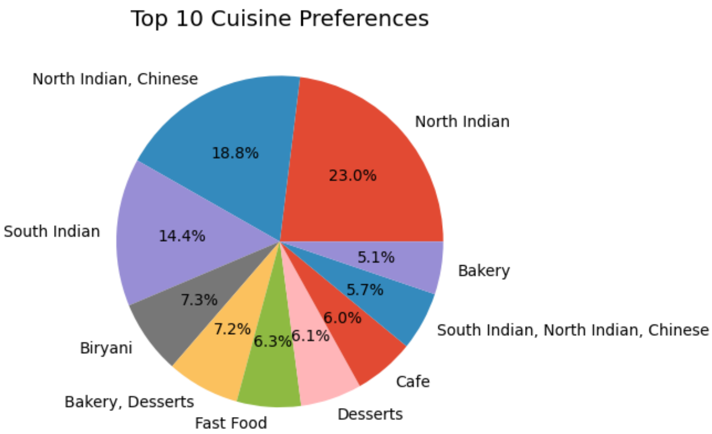

Fig 3.3 Pie chart plotting top 10 cuisine preferences

VIII. CONCLUSION

In conclusion, the Zomato prediction of ratings, sentiment analysis, and recommendation system is a helpful tool for users to make educated judgements about where to eat and for the business and tie-ups to analyse the performance. Users may identify restaurants that meet their interests for cuisine and price range thanks to the recommendation system and prediction model, which both employ machine learning to anticipate the sentiment of user reviews.

This research focuses primarily on using XAI tools to identify the variables influencing model behaviour. This can provide a deeper understanding of the model's functionality, which is the main goal of this implementation.

IX. FUTURE SCOPE

8. More XAI tools like Seldon Alibi, and interpretML can be used for the reasoning behind model's behavior.
9. Different ML algorithms can be implemented to enhance the accuracy and experience for the users.
10. These analysis models can be tried on other regions and can be expanded.

REFERENCES

[1] R. S. Jagdale and S. S. Deshmukh, "Sentiment Classification on Twitter and Zomato Dataset Using Supervised Learning Algorithms," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), Aurangabad, India, 2020, pp. 330-334, doi: 10.1109/ICSIDEMPC 49020.2020.9299582.

[2] S. Lee, H. Shin, I. Choi and J. Kim, "Analyzing the Impact of Components of Yelp.com on Recommender System Performance: Case of Austin," in IEEE Access, vol. 10, pp. 128066-128076, 2022, doi: 10.1109/ACCESS.2022. 3225190.

[3] Aye, Yu Mon, and Sint Sint Aung. "Sentiment analysis for reviews of restaurants in Myanmar text." In 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/ Distributed Computing (SNPD), pp. 321-326. IEEE, 2017.

[4] Hossain, Naimul, Md Rafiuzzaman Bhuiyan, Zerin Nasrin Tumpa, and Syed Akhter Hossain. "Sentiment analysis of restaurant reviews using combined CNN-LSTM." In 2020 11th International conference on computing, communication and networking technologies (ICCCNT), pp. 1-5. IEEE, 2020

[5] K. Zahoor, N. Z. Bawany and S. Hamid, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning," *2020 21st International Arab Conference on Information Technology (ACIT)*, Giza, Egypt, 2020, pp. 1-6, doi: 10.1109/ACIT50332.2020.9300098.

[6] Obiedat, Ruba, Raneem Qaddoura, Al-Zoubi Ala'M, Laila Al-Qaisi, Osama Harfoushi, Mo'ath Alrefai, and Hossam Faris. "Sentiment analysis of customers' reviews using a hybrid evolutionary svm-based approach in an imbalanced data distribution." *IEEE Access* 10 (2022): 22260-22273.

[7] De Albornoz, Jorge Carrillo, Laura Plaza, Pablo Gervás, and Alberto Díaz. "A joint model of feature mining and sentiment analysis for product review rating." In *Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings 33*, pp. 55-66. Springer Berlin Heidelberg, 2011.

[8] Gao, Yifan, Wenzhe Yu, Pingfu Chao, Rong Zhang, Aoying Zhou, and Xiaoyan Yang. "A restaurant recommendation system by analyzing ratings and aspects in reviews." In *Database Systems for Advanced Applications: 20th International Conference, DASFAA 2015, Hanoi, Vietnam, April 20-23, 2015, Proceedings, Part II 20*, pp. 526-530. Springer International Publishing, 2015.

[9] IŞIK, MUHİTTİN, and Hasan Dağ. "The impact of text preprocessing on the prediction of review ratings." *Turkish Journal of Electrical Engineering and Computer Sciences* 28, no. 3 (2020): 1405-1421.

[10] Ganu, Gayatree, Noemie Elhadad, and Amélie Marian. "Beyond the stars: improving rating predictions using review text content." In *WebDB*, vol. 9, pp. 1-6. 2009.

[11] Gunawardena, Dinushika, and Kumuduni Sarathchandra. "Bestdish: A digital menu and food item recommendation system for restaurants in the hotel sector." In *2020 International Conference on Image Processing and Robotics (ICIP)*, pp. 1-7. IEEE, 2020.

[12] 7. Krishnaraj, Sathyaseelan. (2021). Restaurant Recommendation System Using Machine Learning Algorithms. RIET-IJSET International Journal of Science Engineering and Technology. 9.

[13] Sawant, Sumedh, and Gina Pai. "Yelp food recommendation system." *SawantPai-YelpFoodRecommendationSystem. pdf* (2013).

[14] Poriya, Anil, Tanvi Bhagat, Neev Patel, and Rekha Sharma. "Non-personalized recommender systems and user-based collaborative recommender systems." *Int. J. Appl. Inf. Syst* 6, no. 9 (2014): 22-27.

[15] Gupta, Anant, and Kuldeep Singh. "Location based personalized restaurant recommendation system for mobile environments." In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 507-511. IEEE, 2013.

[16] Ashok, Meghana, Swathi Rajanna, Pradnyesh Vineet Joshi, and Sowmya Kamath. "A personalized recommender system using machine learning based sentiment analysis over social data." In *2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1-6. IEEE, 2016.

[17] Zeng, Jun, Feng Li, Haiyang Liu, Junhao Wen, and Sachio Hirokawa. "A restaurant recommender system based on user preference and location in mobile environment." In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 55-60. IEEE, 2016.

[18] R. Katarya and O. P. Verma, "Restaurant recommender system based on psychographic and demographic factors in mobile environment," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Greater Noida, India, 2015, pp. 907-912, doi: 10.1109/ICGCIoT.2015.7380592.

[19] Jonathan, Bern, Jay Idoan Sihotang, and Stanley Martin. 2019. “Sentiment Analysis of Customer Reviews in Zomato Bangalore Restaurants Using Random Forest Classifier”. *Abstract Proceedings International Scholars Conference* 7 (1):1719-28. https://doi.org/10.35974/isc.v7i1.1003.

[20] Bhoite, Sachin & Kulkarni, Atharva & Bhandari, Divya. (2019). Restaurants Rating Prediction using Machine Learning Algorithms. International Journal of Computer Applications Technology and Research. 8. 377-378. 10.7753/IJCATR0809.1008.

[21] Zahoor, Kanwal & Bawany, Narmeen & Hamid, Soomaiya. (2020). Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning. 1-6. 10.1109/ACIT50332.2020.9300098.

[22] Deep Kalbande , Pushpak Patil , Sahil Kale , Siddhant Kasar , Sanjeevkumar Angadi , Prasad Dhore. 2022. “A Survey of Sentimental Analysis on Zomato Restaurant Reviews”. *Journal of Pharmaceutical Negative Results*, December, 7549-56. https://doi.org/10.47750/pnr.2022.13.S07.910.

[23] Arpit Saxena , Prof. Nidhi Sengar , Prof. Amita Goel, Prof. Vasudha Bahl4, “Zomato Data Analysis”. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; (https://doi.org/ 10.22214/ijraset.2021.39303)