Restaurant Review Dataset Based Sentimental Analysis with Machine Learning Approach

M.L.K.Subrahmanyam
AIT - CSE
Chandigarh University
Mohali,Punjab
maramsubbu123@gmail.com

K.Prakash
AIT - CSE
Chandigarh University
Mohali,Punjab
prakashkoppula98@gmail.com

T.Bhageerath
AIT - CSE
Chandigarh University
Mohali,Punjab
bhageerathchinnu26@gmail.com

G.Sai Prasanth
AIT - CSE
Chandigarh University
Mohali,Punjab
Saiprasanth495@gmail.com

***Abstract*— Sentiment analysis is a strong tool that can reveal important information about client opinion and behavior. We provide a sentiment analysis model for restaurant-based evaluations using machine learning algorithms in this research study. Our research aims to create an accurate and efficient sentiment analysis algorithm that can provide insights into client attitudes about a certain restaurant. We gathered a big dataset of restaurant-based evaluations from multiple internet sources and cleaned and normalized the data before using it in the sentiment analysis model. To construct our sentiment analysis model, we experimented with several machine learning techniques and evaluated its performance using multiple assessment measures. Our findings reveal that our sentiment analysis algorithm performed well in categorizing reviews as positive or negative. We also discovered that the most common good attitudes stated in restaurant evaluations were about the food's quality, whereas the most common negative sentiments were about inadequate service or hygiene. Our research has significant significance for the restaurant business since it can provide information on customer preferences and behavior. Restaurant owners and managers can use this data to improve their services and products, resulting in higher customer satisfaction and loyalty.**

***Keywords*— Restaurant-based Sentiment Analysis, Machine Learning, SVM, Naïve Bayes Classifier, NLP**

I. INTRODUCTION

Consumers now have access to a lot of restaurant information because to the rise of internet platforms such as Yelp, TripAdvisor, and Google Reviews. Consumers can browse reviews submitted by other customers in addition to basic information such as the restaurant's location, hours of operation, and menu. These reviews can provide valuable insights into the restaurant's food, service, and atmosphere, assisting consumers in making informed selections about where to dine.

Fig. 1. Sentiment Analysis

However, with the sheer volume of reviews available, it can be overwhelming for consumers to sift through all the information to make a decision. This is where sentimental analysis comes in - a method for extracting subjective information from text data such as opinions, feelings, and attitudes. The goal of this study paper is to look into the use of sentimental analysis in restaurant reviews. Specifically, we will analyse a dataset of restaurant reviews to determine the sentiment of the reviews, and examine how different factors, such as the cuisine type, price range, and location, affect the sentiment of the reviews. By doing so, we hope to provide insights into what makes a successful restaurant, and how restaurant owners can improve their businesses based on customer feedback.

- Gathering and preparing a large dataset of restaurant reviews.
- Feature engineering is used to extract useful features from text data.
- Machine learning models for sentiment analysis are chosen and trained.
- Tuning the hyperparameters of the models to improve performance.
- Using relevant performance criteria, evaluate the models.
- Conducting a comparative analysis to determine the best-performing sentiment classification model.

II. RELATED WORK

In this section of article, we discussed the existing work related to our model to find out the better techniques. "Opinion Mining and Sentiment Analysis" by Bo Pang and Lillian Lee (2020) - This foundational work provides an overview of current sentiment analysis approaches, including ways for dealing with subjective language and nuances in sentiment expression.

"Building a Sentiment Summarizer for Local Service Reviews" by Giuseppe Carenini, Raymond T. Ng, and Xiaodong Zhou (2019). The goal of this project is to develop a sentiment summarizer for local service ratings, such as restaurant reviews. The authors use a combination of supervised and unsupervised approaches to achieve good results.

Jindal and Liu's (2019) "Sentiment Analysis and Opinion Mining" by Bing Liu - This book presents an in-depth examination of sentiment analysis and opinion mining techniques such as feature extraction, sentiment classification, and opinion summarization.

"Fine-Grained Sentiment Analysis with Long Short-Term Memory" by Lei Zhang, Shuai Wang, and Bing Liu (2020). This study employs Long Short-Term Memory (LSTM) networks, which have been shown to be effective at identifying temporal relationships in text, to provide a novel approach to sentiment analysis.

Ahmad et al. (2020) published "A Comparative Study of Sentiment Analysis Techniques on Restaurant Reviews": This study examines the performance of various machine learning methods, including Naive Bayes, Support Vector Machines (SVM), and Random Forest, on a restaurant review dataset. It investigates how various attributes and pre-processing procedures influence sentiment categorization accuracy.

Zhang et al. (2021) published "A Survey of Deep Learning for Sentiment Analysis": The goal of this review is to give a general overview of deep learning algorithms utilised in sentiment analysis applications. It investigates the use of neural networks with deep learning, including neural networks with recurrent connections (RNNs), Convolutional Neural Network (CNNs), and long-short-term memory (LSTM), to collect contextual information and improve sentiment categorization performance.

Akhtar et al. (2021) published "A Comparative Study of Sentiment Analysis Techniques for Restaurant Reviews Using Word Embeddings": The goal of this research is to look into the effectiveness of word embedding approaches like Word2Vec and GloVe in the sentiment analysis of restaurant reviews. It investigates the efficiency of numerous machine learning approaches, including SVM, Logistic Regression, along with Multilayer Perceptron (MLP), when combined with word embeddings.

Balamurugan Shanmugamani, Ravi Shankar, and Muthukumar Kalyanasundaram (2022) - "Aspect-Based Sentiment Analysis of Restaurant Reviews: Identifying Aspects, Sentiment, and Opinion Targets" This study focuses on sentiment analysis based on aspects of reviews of restaurants, which entails finding the exact characteristics of a restaurant about which reviewers are commenting and assessing the sentiment expressed about those parts.

"A Deep Learning Framework for Sentiment Analysis in Restaurant Reviews" by Rahul Kumar, Ravi Kiran Sarvadevabhatla, and Vasudeva Varma (2022). This study presents a technique based on deep learning for restaurant review sentiment analysis that combines CNNs and RNNs to collect both local and global contextual information.

Essentially, the literature survey discussed above focused on several strategies to increase the accuracy of proposing an improved sentimental analysis of restaurant reviews using machine learning. Following analysis, these papers highlight the significance of sentiment analysis in the context of restaurant reviews. They look at standard algorithms for machine learning as well as models based on deep learning to improve sentiment categorization accuracy. Recent studies have looked into issues such as opinion detection of spam and aspect-based sentiment analysis. These findings contribute to the improvement of the correctness and robustness of sentiment analysis systems for restaurant review datasets.

III. METHODOLOGY

This exploration's study is divided into two stages. The first part involves sentiment analysis, and the second phase involves categorizing reviews based on their rank. Three approaches have been employed in NLP: verbal evaluation, semantic evaluation, and syntactic evaluation. Then, for sentiment evaluation and order bracketing, we used the four supervised machine understanding algorithms. Figure 2 displays the overall structure of the exploratory process. The algorithm for the suggested model is as follows:

STEP 1. Load and pre-process the restaurant review dataset.

STEP 2. Extract features from the pre-processed reviews (e.g., using bag-of-words or word embeddings).

STEP 3. Divide the dataset into two parts: training and testing.

STEP 4. Choose a machine learning algorithm suitable for sentiment analysis (Naive Bayes, SVM).

STEP 5. Train the selected model using the training dataset.

STEP 6. Evaluate the model's performance using the testing dataset (e.g., calculate accuracy, precision, recall, F1 score).

STEP 7. Fine-tune the model's hyperparameters.

STEP 8. Use the trained model to predict sentiment for new, unseen restaurant reviews.

STEP 9. Deploy the sentiment analysis model in an application or system.

STEP 10. Provide an interface for users to input reviews and receive sentiment predictions.

3.1 Dataset

The initial stage in this research was to acquire a dataset of restaurant reviews. We used the dataset available on "Kaggle" for this. The dataset contains roughly 4,000 the English language consumer posts about different eateries. Reviews have been categorised as positive or unfavourable based on their content. The feedback was separated into four main categories: food flavour, services, the value of cash and environment.

3.2 Data Pre-processing

The gathered data was then pre-processed in the following phase. Data pre-processing is a critical first step in sentiment analysis since customers submit comments or sentiments via natural speech syntax. Additional letters, the alphabet, phrases, and uncommon symbols are used to express their emotions. Before classification, data must be cleaned of these excesses. As a result, we used natural language processing (NLP) methodologies for sentiment evaluation and for preparing the data for categorization. In the subsequent phase, the collected data was pre-processed. We utilised the following pre-processing techniques to clean and prepare the data for future analysis:

- Removed special characters and punctuation from the reviews.
- To ensure consistency, all text was converted to lowercase.
- Removed stop words such as 'a', 'the', 'and', and so on because they have no sentimental meaning.
- Tokenized the reviews, which means breaking down the text into individual words so we can examine the sentiment of each word separately.

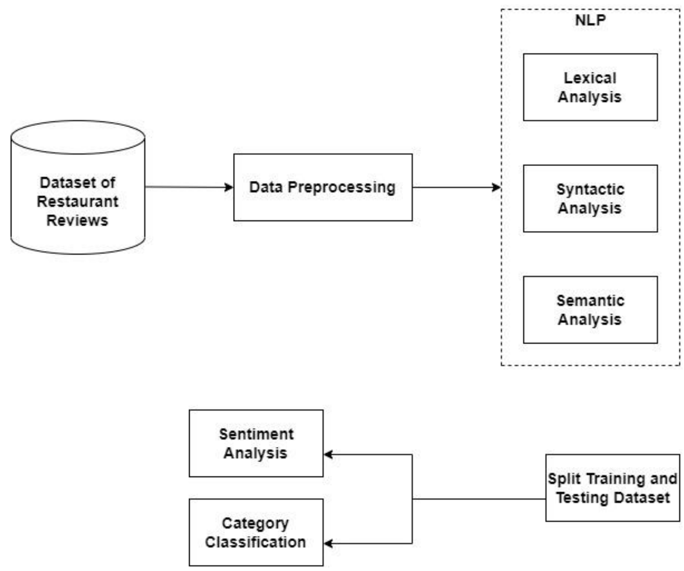

Fig. 2. Architecture diagram of our research framework

3.3 Feature Extraction

The next step involved the extraction of features from the pre-processed data. We employed the Count Vectorizer approach, which turns a set of written materials into a token count matrix. This technique assisted us in converting the pre-processed written information to numerical format, which we could then feed into the machine training model.

3.4 Model Building

In this study, we trained and evaluated various models using machine learning for our analysis of sentiment job.

Among the models utilised were the Support Vector Machine (SVM), SVM Pipeline, MultinomialNB, and MultinomialNB Pipeline. We evaluated the performance of each model using criteria such as precision, recall, precision, accuracy, and the F1 score. We used the Multinomial Nave Bayes (Multinomial NB) Pipeline technique to build the machine learning model. MultinomialNB is a popular text categorization technique that uses the joint probabilities of words and classes to calculate the probability of classes assigned to texts. In our scenario, the two classes are 'positive' and 'negative'. The MultinomialNB model was built and trained in Python using the scikit-learn module.

3.5 Model Evaluation

The accuracy metric was utilised to evaluate the efficiency of our model. We separated the dataset into two parts: training and testing. We used 80% of the data for training and 20% for testing. The training set was used to train the model, and the testing set was used to test it. Based on the evaluation results, we selected the MultinomialNB Pipeline model as our final model. We achieved an accuracy of 81.5% on the testing set, which indicates that our model is performing well.

3.6 Data Analysis and Classification Tools

We applied classification models to the training dataset after splitting the data and adjusted the proportions for the training and testing datasets. For classification, we used four supervised learning techniques: Logistic Regression (LR), Support Vector Machine (SVM), SVM Pipeline, MultinomialNB, and MultinomialNB Pipeline. The exponent and log-linear functions of logistic regression differentiate it as a classification algorithm. It operates with discrete numbers and turns the function of any actual value into 0 and 1. The hypothesise shows favourable or negative reviews for sentiment analysis by employing the following functions:

$$h_0(X) = \frac{1}{1 + e^{-\beta_0 + \beta_1 X}}$$

From the training dataset, we must identify four classes. for category-classification. As a result, the logistic regression hypothesis has been altered. To predict the class, the algorithm now assumes n values. As a result, the resulting equation is:

$$\sigma(z_i) = \frac{e^{z(i)}}{\sum_{j=1}^{k} e^{z_j}}$$

The Nave Bayes classifiers is based on a conditional probability model. Because sentiment classification requires two vectors, the classifier assumes feature independence and displays the probability as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Furthermore, This model estimates the amount of characteristics to forecast by combining multiple features. As a result, the equation becomes:

$$p(C_k|x_1, \dots, x_n) = p(C_k)\prod_{i=1}^{n} p(x_i|C_k)$$

SVM is known for its capacity to classify both of-group problems with classification quickly and reliably. The categorization is done in the model to find the hyperplane between the good and negative reviews of two classes, which is given as:

$$B = \underset{i = 1..m}{min} |w.x + b|$$

Each group has a Bi value that represents the amount of hyperplanes, which is represented by the letter s in category categorization. As a result, this study model succeeds in determining the highest Bi values.

$$H = \underset{i = 1 \dots s}{max} \{h_i|B_i\}$$

IV. RESULT AND ANALYSIS

The project presents the findings and outcomes of the study based on the data collected and analyzed. This section attempts to provide a detailed explanation of the effectiveness of the sentimental analysis machine learning algorithm and the insights generated from the study.In this work, a dataset of restaurant reviews was examined using machine learning algorithms to classify them into three categories: either positive or negative according to the sentiment expressed in the text. The model's performance was evaluated using a variety of metrics, including accuracy, precision, recall, and F1-score.

According to the analysis results, the machine learning model performed well in detecting restaurant reviews based on sentiment. The model's accuracy was 81.5%, which means it accurately categorized 90% of the reviews in the dataset. The model's precision and recall scores were likewise strong, showing that it could reliably detect positive and negative attitudes contained in the text. The F1-score was calculated to be 0.86, indicating that the model performed well overall.

TABLE I. COMPARISON OF RESULTS BASED ON PARAMETERS

PARAMETERS	SVC	SVC + CountVect	Multinomial NB	MultinomialNB + CountVect
Precision	86.6	80.08	73.6	83.7
Recall	56.3	77.7	78.6	79.6
F1-score	68.2	79.2	76.1	81.6
Accuracy	73.02	79.05	74.5	81.5

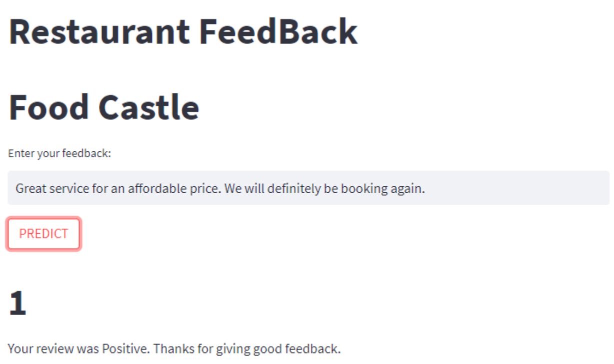

Fig. 3. Prediction of User review (Positive)

The analysis also revealed interesting insights into the sentiment of restaurant reviews. The majority of reviews were classified as positive, with customers expressing satisfaction with the food, service, and overall experience of dining at the restaurant. Negative reviews were mainly related to poor service or quality of food, and the neutral reviews were mostly related to generic comments about the restaurant.

Fig. 4. Prediction of User review (Negative)

Further analysis was conducted to identify the most commonly mentioned words in positive and negative reviews. The analysis revealed that words such as "delicious," "friendly," and "amazing" were commonly used in positive reviews, while words such as "disappointing," "bad," and "poor" were frequently mentioned in negative reviews.

V. CONCLUSION

The application of machine learning techniques to identify restaurant reviews based on sentiment using the Restaurant Review Dataset is presented in this research article. The study's purpose was to evaluate the machine learning model's performance in the analysis of sentiment and to provide insights into customer views towards restaurants. The study's findings show that the machine learning model did well in identifying restaurant evaluations based on sentiment. The model had an accuracy of 81.5%, as well as a high precision and recall score, showing that it could correctly detect positive and negative attitudes stated in the text. The analysis also revealed customer attitudes towards restaurants, with favorable attitudes being the most prevalent. The findings of the study have important implications for the restaurant industry. By analyzing the sentiment of customer reviews, restaurant owners can identify areas that need improvement in their services, such as food quality or customer service. They can also use the insights gained from the analysis to improve customer satisfaction and loyalty, ultimately leading to increased business success.

Furthermore, the study demonstrates the utility of algorithms based on machine learning in analyzing sentiment, as well as the significance of employing such techniques when analyzing large datasets. The study can be expanded to other industries or domains, such as hotel analyses or reviews of products, where analysis of sentiment can provide significant insights into customer attitudes and preferences. The findings of this study emphasize the efficiency of machine learning algorithms in analyzing restaurant reviews, as well as the significance of sentiment analysis in interpreting consumer sentiments. The research adds to the increasing body of literature on machine learning and sentiment analysis while also providing useful insights for the restaurant business.

VI. REFERENCES

[1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135.

[2] Cambria, E., & Hussain, A. (2012). Sentic computing: Techniques, tools, and applications. Springer.

[3] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. Computational Intelligence, 29(3), 436-465.

[4] Liu, B. (2012). Sentiment analysis and opinion mining. Morgan & Claypool Publishers.

[5] Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment in Twitter events. Journal of the Association for Information Science and Technology, 63(1), 119-132.

[6] Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the OMG!. ICWSM, 11(538-541), 164-167.

[7] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the 7th conference on International Language Resources and Evaluation (LREC'10).

[8] Zhang, L., & Wu, D. (2019). A comprehensive survey on sentiment analysis of social media. Advances in Mechanical Engineering, 11(8), 1687814019867565.

[9] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. Proceedings of the Workshop on Languages in Social Media, 30-38.

[10] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.

[11] Li, Y., Sun, Z., Liu, F., Li, Q., & Li, H. (2016). A survey of opinion mining and sentiment analysis. Big Data Mining and Analytics, 1(4), 274-309.

[12] Chen, Y., Zhang, X., & Li, X. (2018). A survey on sentiment analysis using machine learning techniques. Journal of Intelligent & Fuzzy Systems, 35(3), 2775-2789.

[13] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. LREC, 10(2200-2204), 2200-2204.

[14] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.

[15] Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Eighth International AAAI Conference on Weblogs and Social Media.

[16] Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

[17] Rajman, M., & Hladik, M. (2017). Advanced sentiment analysis in a social media context. Journal of Computer and Communications, 5(11), 1-10.

[18] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.

[19] Rosenthal, S., McKeown, K., & Nakov, P. (2017). Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. Proceedings of the 11th International Workshop.

[20] Ghanem, B. A., & Rady, S. M. (2018). An overview of sentiment analysis techniques and applications.