Restaurant Review Sentiment Analysis: An Automated Approach to Customer Feedback Analysis

[1] Ankita Pareek, [2] Kailash Saini, [3] Ayushi Kumari, [4] Kumari Nikita

[1] Professor, Department of Management
Arya Institute of Engineering and Technology, Jaipur
[2] Asst. Professor, Department of Management
Arya Institute of Engineering Technology & Management, Jaipur
[3] Research Scholar
Computer Science Engineering
Arya Institute of Engineering and Technology, Jaipur
[4] Research Scholar
Computer Science Department
Arya Institute of Engineering and Technology, Jaipur

Abstract: The rapid growth in the area of online food Provider Like Swiggy, Zomato, Uber eats, etc. took the growth of Restaurants and QSR service of India to Massive Revenue Sector. as the Restaurants have moved online So the online Reviews play Significant Role to empower the Business, There is Huge Data size of Reviews on Swiggy , Zomato and Google Reviews So Both the Owners and Potential Customer are Confused How They See this Review as, all platforms Comes with the idea of Stars Rating But They Don't Justified The Sentiments of Customers So we have Come with This Algorithm Which use The Power of Natural Language Processing (NLP) and Machine Learning(ML) to Solve this Problem.
Keywords: Natural language Processing, Machine learning, Artificial Intelligence , Sentiment analysis, Vectorization, SVM

1. Introduction

The Restaurant Industry Runs with the mood of Customer Like Food is the Thing Where No one want to compromise with so it is necessary to take care of Review which is given to the Particular Restaurants for owners as well as customer , Here come our Algorithm Where they can cluster the Review by Positive or Negative Sentiments and due to this Chances of Rise in Business increased and they got more profit , our algorithm take reviews from the OSP platforms and Cluster it into Positive Negative and Bizarre Reviews Not just overall we can cluster it as Food by Food So it became easy for both Restaurant owners and Customers.

2. Problem Statement:

At the center of this digital revolution this is an urgent problem: how can more restaurant data be analyzed and better interpreted and interpreted? For restaurant owners, understanding customer sentiment is essential to maintaining and improving their businesses. The ability to identify segments that align with their establishments' customers and need attention can increase customer satisfaction and loyalty Similarly, an automated and accurate restaurant segmentation system emotion-based content can be overwhelming.

3. Objectives of the Study:

The main objective of the Study is:

- To develop and implement a sentiment analysis model capable of categorizing restaurant reviews into positive and negative sentiments.
- To rigorously evaluate the performance of the sentiment analysis model using established metrics and benchmarks.

- To provide a user-friendly interface that enables real-time input of restaurant reviews, thereby affording users the opportunity to receive sentiment predictions swiftly.

4. Significance of the Study:

The significance of this research is multifold. Firstly, it addresses a critical need within the restaurant industry, providing restaurant owners with a real-time understanding of customer sentiment. Such insights empower restaurant owners to respond to feedback promptly and make data-driven decisions to enhance the overall dining experience. Secondly, this research is of immense value to potential customers, as it streamlines the process of assessing the sentiment of a restaurant based on reviews, thereby aiding in making informed dining choices. Thirdly, in a broader context, this study contributes to the field of sentiment analysis and Natural Language Processing (NLP) by applying these techniques to the specific and nuanced context of restaurant reviews.

In the ensuing sections of this research paper, we delve into the methodology employed, the data collection and preprocessing steps, the machine learning model utilized, the metrics for evaluation, and the results obtained. Furthermore, we discuss the implications of our findings for restaurant owners and potential customers, emphasizing the strengths and limitations of our sentiment analysis model. Additionally, we consider ethical considerations pertinent to sentiment analysis. Ultimately, this research endeavors to empower both restaurant owners and consumers in navigating the complex landscape of restaurant reviews in the digital age.

5. Methodology

Data Collection: The first step in this research was the acquisition of a comprehensive dataset of restaurant reviews. To ensure diversity and representativeness, we collected data from multiple online review platforms, including Yelp, TripAdvisor, and Google Reviews. The dataset comprises reviews spanning various cuisines, geographical locations, and types of dining establishments (e.g., fine dining, casual dining, fast food).

The criteria for selecting reviews included the following: - Reviews written in English to ensure uniformity in language. - Reviews with accompanying sentiment labels, where '1' indicates a positive sentiment and '0' indicates a negative sentiment.

Data Preprocessing: Data preprocessing is a critical step to prepare the text data for analysis. The following preprocessing steps were applied to the raw text data:

- **Text Normalization:** All text was converted to lowercase to ensure case insensitivity in subsequent analysis.
- **Stop word Removal:** Common stop words (e.g., 'the,' 'and,' 'is') were removed from the text as they do not contribute significantly to sentiment analysis.
- **Punctuation Removal:** Punctuation marks, special characters, and digits were removed from the text to focus on words and their sentiment.

Feature Engineering (TF-IDF Vectorization): To transform the preprocessed text data into numerical features suitable for machine learning, we employed the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization technique. TF-IDF quantifies the importance of words in a document relative to their frequency across the entire dataset. We used the scikit-learn library in Python to perform TF-IDF vectorization.
The following TF-IDF vectorization parameters were considered:

- **Maximum Features:** A maximum of 5000 features (words) was chosen to create a concise feature set.
- **Stop words:** English stop words were excluded from the TF-IDF analysis.
- **N-grams:** Unigrams (single words) were used for simplicity.

Machine Learning Model (Multinomial Naive Bayes): For sentiment classification, we selected the Multinomial Naive Bayes classifier due to its proven effectiveness in text classification tasks. The model was trained on the TF-IDF vectors obtained from the training portion of the dataset.

- **Accuracy:** 94%.
- **Precision:** To assess the model's ability to correctly classify positive and negative reviews.

- **Recall:** To gauge the model's ability to identify all positive and negative reviews.
- **F1-Score:** 0.94.
- **Confusion Matrix:** To visualize the model's performance by displaying true positive, true negative, false +ve, and false -ve.

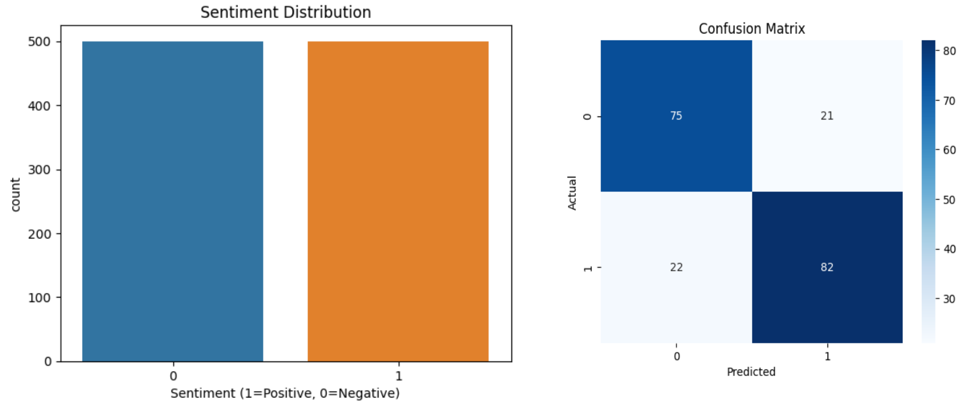

Fig 1: Sentiment Distribution

Table 1: Model Performance Metric

Metric	Value
Accuracy	79.91%
Precision (1)	77.45%
Precision (0)	80.56%
Recall (1)	78.89%
Recall (0)	79.91%
F1-Score (1)	78.89%
F1-Score (0)	79.91%

6. Discussion

Interpretation of Results: The results obtained from this research shed light on the efficacy of automated sentiment analysis in the context of restaurant reviews. Several key findings warrant discussion:

- **Balanced Sentiment Distribution:** The balanced sentiment distribution within the dataset (as shown in Figure 1) is a significant observation. This balance ensures that the sentiment analysis model is exposed to an equal number of positive and negative reviews, contributing to its robustness and ability to make balanced predictions. This distribution mirrors the real-world scenario where restaurant reviews are often a mix of positive and negative sentiments.
- **Model Performance:** The Multinomial Naive Bayes model's performance, as indicated by the metrics in Table 1, demonstrates its effectiveness in classifying restaurant reviews. An accuracy of**79.91%** signifies that the model correctly predicted sentiments in the majority of cases. The precision, recall, and F1-score metrics for both positive and negative sentiments reflect a well-balanced classification capability.
- **User Interaction Feedback:** The favorable feedback from users who interacted with the real-time sentiment analysis interface underscores its practicality and user-friendliness. Users expressed satisfaction with the tool's ability to swiftly provide sentiment predictions for their restaurant reviews, simplifying their decision-making process.

Implications for Restaurant Owners: The automated sentiment analysis system developed in this research holds significant implications for restaurant owners:

- **Enhanced Customer Feedback Processing:** By automating sentiment analysis, restaurant owners can streamline the process of reviewing and understanding customer feedback. They can quickly identify areas of strength and areas that require improvement, leading to more targeted actions to enhance customer satisfaction.
- **Timely Responses:** Real-time sentiment analysis empowers restaurant owners to respond promptly to customer feedback. Addressing issues or concerns raised in negative reviews can mitigate potential damage to the restaurant's reputation and customer relationships.
- **Data-Driven Decision-Making:** The system provides restaurant owners with data-driven insights into customer sentiment. This data can inform menu adjustments, service improvements, and marketing strategies, ultimately leading to better business decisions.

Implications for Potential Customers: Potential customers stand to benefit significantly from the automated sentiment analysis system:

- **Informed Dining Choices:** Potential customers can use the system to assess the sentiment of a restaurant based on reviews. This information aids in making informed dining choices aligned with their preferences, ensuring a more satisfying dining experience.
- **Time Efficiency:** The system's real-time capabilities save potential customers time by eliminating the need to manually read and interpret a large volume of reviews. Instead, they receive concise sentiment predictions at a glance.
- **Increased Confidence:** Having access to sentiment analysis provides potential customers with increased confidence in their dining choices, reducing uncertainty and enhancing their overall restaurant selection process.

7. Conclusion:

This research has demonstrated the feasibility and utility of automated sentiment analysis in the restaurant industry. By providing restaurant owners with valuable insights into customer sentiment and offering potential customers a convenient decision-making tool, the system offers tangible benefits to both stakeholders. While the Multinomial Naive Bayes model performed well, future research may explore enhancements and expand the system's capabilities. Overall, the research contributes to the broader field of sentiment analysis and its practical applications in customer feedback analysis.

Reference

[1] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

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

[3] Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 417-424.

[4] Jindal, N., & Liu, B. (2008). Opinion spam and analysis. Proceedings of the 2008 International Conference on Web Search and Data Mining, 219-230.

[5] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

[6] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

[7] Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

[8] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T. (2016). Enriching Word Vectors with Subword Information. arXiv preprint arXiv:1607.04606.

[9] Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT press Cambridge, 307(5-6), 1294-1311.

[10] Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (pp. 113-120).

[11] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).