**Sentiment Analysis of Restaurant Reviews Using Random Forest Model**

**Chris Angelu Bongcawil Jordan1, Mike Rassel Carale Dagooc2**

Computer Science and Information Technology Department,

Bachelor of Science in Computer Science (BSCS)

Negros Oriental State University (NORSU), Negros Oriental, Philippines

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Article history:***  Received November 14, 2024  Revised November 14, 2024  Accepted January 15, 2025 |  | This study employed a Random Forest classifier to perform sentiment analysis on 1,502 restaurant reviews across Europe, utilizing a combination of textual data, such as customer reviews, and associated metadata, to enhance classification accuracy. The preprocessing pipeline included text cleaning, stop-word removal, TF-IDF vectorization, and feature selection. The objective was to develop a robust model capable of categorizing reviews into positive and negative sentiments while identifying key factors influencing customer satisfaction. The Random Forest model achieved an overall accuracy of 92%, with 0.93 precision and 0.92 recall values demonstrating its reliability and effectiveness. Further analysis revealed that negative reviews commonly mentioned issues like poor service, overpriced food, and substandard cleanliness, whereas positive reviews frequently highlighted exceptional food quality, ambiance, and friendly staff. Word clouds and feature importance visualizations provided insights into recurring themes in customer feedback. The findings underscore the Random Forest model’s utility as a practical and interpretable tool for real-time sentiment analysis, offering actionable recommendations for improving restaurant services and enhancing customer satisfaction. |
| ***Keywords:***  *Sentiment Analysis*  *Random Forest*  *Machine Learning*  *Consumer Feedback* |
|  |

1. **INTRODUCTION**

Customer satisfaction, a cornerstone of customer loyalty and business success, holds particular importance in the highly competitive restaurant industry. With dining establishments serving millions of customers daily across Europe, feedback from patrons provides valuable insights into the quality of services and overall experience [1], [3]. Sentiment analysis of such reviews can help restaurants better understand customer perceptions, address service gaps, and improve satisfaction. However, manually analyzing a large volume of informal and subjective reviews is inefficient and prone to bias or error [4], [5], [6].

Although advanced deep learning models offer high accuracy, they often require significant computational resources and lack transparency, making them less suitable for practical applications in real-time scenarios. Alternatively, traditional machine learning models, such as Random Forest classifiers, strike a balance between accuracy, resource efficiency, and interpretability, offering practical solutions for sentiment classification tasks [2], [7].

This study contributes to the field of sentiment analysis by developing a tailored Random Forest model to classify restaurant reviews into positive and negative sentiments. Unlike previous studies that focus solely on textual data, this research leverages a combination of textual reviews and associated metadata, such as numerical ratings, to enhance classification accuracy and derive comprehensive insights into customer satisfaction [8], [9], [10].

Innovative features of the study include the use of word clouds to visualize prominent themes in customer feedback and feature importance analysis to identify critical factors influencing sentiment classification. The results reveal common complaints in negative reviews, such as poor service and overpriced food, alongside recurring praise in positive reviews for qualities like excellent food, ambiance, and friendly staff [12], [13], [14].

The primary objective of this research is to design a model that combines simplicity and effectiveness to analyze customer sentiment and provide actionable insights for the restaurant industry. By leveraging the Random Forest classifier, the study emphasizes practical, transparent, and resource-efficient sentiment analysis, enabling restaurants to adopt data-driven strategies for improving services and enhancing customer satisfaction [17], [18].

1. **LITERATURE REVIEW**

**2.1 Sentiment Analysis**

Sentiment Analysis, also known as opinion mining, is a critical subfield of Natural Language Processing (NLP) that focuses on identifying, interpreting, and categorizing opinions and emotions within textual data. The primary goal is to extract sentiment—positive, neutral, or negative—from data sources like customer reviews, social media posts, and survey responses [19]. This analysis is pivotal in industries such as marketing, customer service, and brand management, enabling organizations to derive actionable insights and make informed decisions based on public opinion [20], [21].

Traditional sentiment analysis approaches include rule-based systems, machine learning models, and statistical techniques. However, recent advancements have introduced deep learning frameworks such as Long Short-Term Memory (LSTM) networks and Bidirectional Encoder Representations from Transformers (BERT), which outperform traditional methods in handling complex linguistic structures. These methods excel in tasks like context-based sentiment prediction, achieving higher accuracy and recall rates [22]. Nonetheless, traditional machine learning methods, such as Random Forest and Naive Bayes, remain widely adopted due to their computational efficiency and simplicity.

In the restaurant industry, customer feedback provides crucial insights into service quality and overall customer experience. Studies have shown that machine learning techniques, including Random Forest and Decision Trees, effectively classify sentiments with high accuracy while being computationally efficient. Although deep learning models outperform in scalability and accuracy, they demand substantial resources and may lack interpretability. Traditional models, such as Random Forest, strike a balance by providing reliable and interpretable results, especially in moderately sized datasets like restaurant reviews [23].

**2.2 Random Forest Classifier**

Random Forest is a robust and flexible machine learning algorithm that operates by constructing multiple decision trees during training and outputting the mode of their predictions in classification tasks or the mean prediction in regression tasks. This ensemble learning method significantly improves classification accuracy while reducing overfitting compared to individual decision trees [24].

Random Forest offers several advantages, including its ability to handle high-dimensional datasets and automatically estimate feature importance. These characteristics make it highly effective for sentiment analysis tasks, where the data often involve complex relationships and numerous input variables. Furthermore, the algorithm's resistance to overfitting ensures reliable performance, even when analyzing noisy data or datasets with missing values.

Despite its strengths, Random Forest has limitations, such as computational overhead in constructing and evaluating numerous trees, which can make it less suitable for real-time applications on massive datasets. However, with optimizations, it remains a popular choice for interpretable and resource-efficient sentiment classification.

In the context of restaurant reviews, Random Forest classifiers have demonstrated high accuracy and reliability, effectively identifying key drivers of customer satisfaction and dissatisfaction. This research leverages Random Forest's strengths to process and classify restaurant reviews, combining textual feedback with numerical features like ratings to derive actionable insights. By highlighting prominent themes in customer feedback through word clouds and analyzing feature importance, this study underscores the practical utility of Random Forest in extracting valuable sentiment-driven insights for improving service quality and enhancing customer satisfaction [25], [26].

1. **METHODOLOGY**

**3.1 Materials**  
 The study utilized a Toshiba laptop with an Intel Core i5-3360m processor and 16GB of RAM, running on Windows 10. This configuration offered adequate computational power to handle the data preprocessing, model training, and evaluation required for sentiment analysis of restaurant reviews.

* + 1. **Software**

The sentiment analysis model was developed using Jupyter Notebook and Python 3.13.0, supported by several Python libraries. Scikit-Learn was utilized to implement the Random Forest algorithm and provide preprocessing utilities, while Pandas and NumPy enabled efficient data manipulation and analysis. Visualization was facilitated by Matplotlib and Seaborn, which provided insightful graphical representations of data trends and model performance. Cloud-based platforms, such as Google Cloud AI, AWS SageMaker, and Microsoft Azure AI, were considered for scalability and deployment, ensuring the robustness of the analytical pipeline.

* 1. **Method**

The dataset used for this analysis was sourced from a repository of restaurant reviews available in CSV format. It contained 1,052 rows and 15 columns, which were sufficient for building a sentiment classification model. The primary objective of the analysis was to categorize customer reviews into positive, negative, or neutral sentiments to gain insights into customer satisfaction with restaurant services. The dataset featured key attributes such as the review text, customer ratings, and attributes like food quality, ambiance, and service. These attributes were utilized to construct a predictive model capable of understanding customer sentiments and drawing actionable insights for improving restaurant operations [27].

* + 1. **Preprocessing**

To ensure the data was suitable for training the model, several preprocessing steps were undertaken. Initially, non-alphabetical characters were removed from the reviews to reduce noise. This was followed by lemmatization to convert words into their base forms and stopword removal to exclude common but uninformative words. Feature scaling was applied to numerical attributes, such as customer ratings for food, service, and ambiance, using the StandardScaler method. Finally, TF-IDF vectorization was used to transform the textual data into numerical features, capturing the importance of individual words and phrases.

* + - 1. **Removing Non-Alphabetical Character**

The first step involved eliminating symbols, numbers, and special characters from the textual data to reduce noise. Reviews often included irrelevant characters such as emojis, hashtags, and numerical expressions that did not contribute to sentiment classification. Removing these extraneous elements ensured the model could focus exclusively on meaningful text, enhancing the overall data quality.

* + - 1. **Lemmatization**

Lemmatization was applied to reduce words to their root forms, ensuring linguistic consistency in the dataset. For instance, words like "eating," "eats," and "ate" were all converted to "eat." This step reduced the dimensionality of the vocabulary, which is critical for sentiment analysis as it aligns similar words under a common representation while preserving their semantic meaning.

* + - 1. **Stopword Removal**

Commonly used words like "the," "is," and "and" were removed during preprocessing. These stopwords, though frequent in text, do not provide meaningful information for sentiment classification. Removing them allowed the model to concentrate on sentiment-rich words, such as "delicious," "poor," "friendly," and "overpriced," which are more indicative of customer sentiment.

* + - 1. **Feature Scaling**

Numerical features, such as ratings for food, ambiance, and service, were standardized to have a mean of zero and a standard deviation of one using the StandardScaler. This step was crucial to ensure that numerical features with varying ranges contributed equally to the model and prevented any single feature from dominating the training process.

* + - 1. **TF-IDF Vectorization**

The final step of preprocessing involved transforming the text data into numerical features using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. This technique assigned weights to words based on their frequency within a specific review and their frequency across the entire dataset. Both unigrams and bigrams were included to capture individual words and meaningful word pairs. For example, a bigram like "poor service" was treated as a single feature, reflecting its strong negative sentiment.

**3.2.2 Word Cloud Visualization**  
 Word clouds were generated to visualize the most frequent terms in positive and negative reviews. For negative reviews, terms such as "rude," "cold," and "overpriced" were highlighted, while positive reviews prominently displayed words like "delicious," "friendly," and "excellent." These visualizations provided an intuitive understanding of the key themes driving customer sentiments.

**3.2.3 Model**

The sentiment analysis model employed was the Random Forest algorithm, known for its robustness and capability to handle large datasets effectively. This ensemble learning method constructs multiple decision trees and combines their predictions to improve accuracy and prevent overfitting. The target variable, customer satisfaction, was encoded as a binary value where 1 represented positive sentiment and 0 represented negative sentiment. This approach allowed the model to classify reviews efficiently and accurately.

**3.2.4 Data Splitting**

The dataset was split into training, validation, and testing subsets in a ratio of 70:15:15, respectively. The training set was used to fit the Random Forest model, while the validation set was employed for hyperparameter tuning, ensuring optimal model performance. The test set provided an unbiased evaluation of the model’s ability to generalize to unseen data.

**3.2.5 Evaluation Metrics**

The model's performance was assessed using multiple metrics, including accuracy, precision, recall, and F1-score. Accuracy provided an overall measure of the model’s correctness, while precision and recall offered insights into its ability to predict positive and negative sentiments. The F1-score, a harmonic mean of precision and recall, was particularly useful in evaluating the model's performance on imbalanced data. Additionally, a confusion matrix was generated to visualize the distribution of true positives, true negatives, false positives, and false negatives, providing a comprehensive understanding of the model’s classification capabilities. Accuracy shows how well the model performs overall by looking at the proportion of correct predictions (both positive and negative) among all predictions.

This metric gives an overall view of how well the model performs across all classes. However, it may not be reliable for imbalanced datasets where one class dominates [27], [28].

(1)

* TP (True Positives): The number of positive instances correctly predicted as positive.
* TN (True Negatives): The number of negative instances correctly predicted as negative.
* FP (False Positives): The number of negative instances incorrectly predicted as positive.
* FN (False Negatives): The number of positive instances incorrectly predicted as negative

Precision focuses on the accuracy of the model’s positive predictions and is important when the cost of false positives is high [27].

(2)

* TP (True Positives): The number of positive instances correctly predicted as positive.
* FP (False Positives): The number of negative instances incorrectly predicted as positive

Recall shows how well the model captures actual positive outcomes by looking at the proportion of correctly identified positives out of all true positives.

Recall is crucial when the cost of false negatives is significant, as it indicates the model's ability to capture positive cases. [28], [29].

* TP (True Positives): The number of positive instances correctly predicted as positive.

(3)

* FN (False Negatives): The number of positive instances incorrectly predicted as negative.

The F1 Score balances Precision and Recall by calculating their harmonic mean. It's particularly useful when you want to account for both false positives and false negatives.

The F1-score is especially useful for imbalanced datasets, as it considers both false positives and false negatives in its calculation [30].

(4)

* Precision: The proportion of true positives out of all predicted positives. Recall:
* The proportion of true positives out of all actual positives.

1. **RESULTS AND DISCUSSION**

#### **4.1 Qualitative Insights from Word Clouds**

 To gain qualitative insights, word clouds were generated to visualize the most frequently occurring terms in both positive and negative reviews. The word cloud for positive reviews highlights terms such as "delicious," "excellent," "friendly," and "service," reflecting customers' appreciation for food quality, staff behavior, and overall dining experience. On the other hand, the word cloud for negative reviews prominently features terms like "rude," "cold," "overpriced," and "disappointing," indicating dissatisfaction with aspects such as service, food temperature, and value for money. These recurring themes provide restaurants with actionable insights to address customer concerns and enhance their offerings.

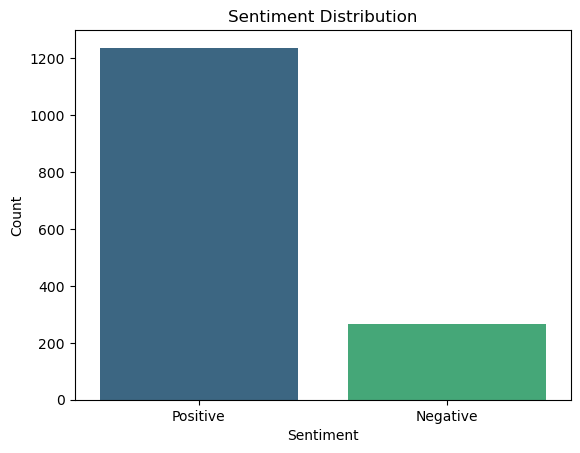
**Figure 1:** Positive Reviews Word Cloud

The most frequently used words in the word cloud include "delicious," "great," "friendly," "amazing," "excellent," and "lovely," which provide a clear reflection of customers' dining experiences. Many reviews highlighted the "food" for its quality and taste, with patrons appreciating the freshness and creativity of the dishes, as noted in one review: "The food was absolutely delicious, and every bite was a treat!" Another customer commented: "The desserts were beautifully presented and tasted even better than they looked." The "service" was another major highlight, with diners praising the friendliness and attentiveness of staff. One review stated: "The waiters were so friendly and accommodating, always ready to make our evening special." Similarly, a guest remarked: "The staff made us feel welcome from the moment we walked in—excellent service all around." The "atmosphere" consistently received commendations for its warmth, charm, and inviting decor. Customers frequently mentioned the ambiance as a key factor in enhancing their overall experience. For instance, a reviewer shared: "The atmosphere was cozy and perfect for a romantic dinner." Another guest noted: "The interior design was stunning, and the lighting created a relaxing vibe." Overall, the word "great" encapsulates the general sentiment of satisfaction, with numerous diners expressing that the restaurant exceeded their expectations in delivering a memorable dining experience. For example, one guest summarized their visit as: "Great food, amazing staff, and an ambiance that made us feel special. Highly recommend!" These aspects collectively highlight the dedication of European restaurants to creating positive and lasting impressions, as reflected in the overwhelmingly positive sentiments in the reviews.

**Figure 2**: Negative Reviews Word Cloud

The most frequently used words in the word cloud include "bad," "poor," "rude," "terrible," "disappointing," and "slow," reflecting customers' frustrations with various aspects of their dining experiences. Many reviews highlighted the "service" as a significant point of dissatisfaction, with patrons expressing disappointment over unresponsive or unfriendly staff. For example, one customer wrote: "The waiter was incredibly rude and seemed uninterested in helping us." Another noted: "Service was extremely slow, and we waited over an hour for our food." The "food" was another frequent source of criticism, with diners mentioning issues such as lackluster taste, poor presentation, and insufficient portions. One reviewer remarked: "The food was cold and tasteless—definitely not worth the price." Similarly, another customer shared: "The steak was overcooked and chewy, and the sides were bland." The "overpriced" was often cited as a concern, with many reviewers feeling that the cost did not match the quality of the experience. A guest complained: "Extremely overpriced for what was essentially a subpar meal." Another review stated: "For these prices, I expected a much higher standard of food and service." The "atmosphere" also received criticism, with customers describing noisy environments, uncomfortable seating, or poor hygiene. For instance, one diner shared: "The restaurant was too loud, making it impossible to have a conversation." Another noted: "The tables were sticky, and the whole place felt unclean." Overall, the word "disappointing" encapsulates the general sentiment of dissatisfaction, with many diners expressing regret over their dining experiences. For example, one reviewer summarized their visit as: "Disappointing from start to finish—bad service, bad food, and an overpriced bill." These recurring themes in negative reviews highlight key areas for improvement, such as enhancing staff training, ensuring food quality and consistency, and addressing concerns about cleanliness and ambiance. By addressing these issues, restaurants can work towards reducing customer dissatisfaction and creating more positive experiences.

* 1. **Sentiment Distribution**



**Figure 3. Sentiment Distribution**

The sentiment distribution provides a visual representation of the proportion of positive and negative sentiments in the dataset of restaurant reviews. A significant majority of the reviews are categorized as positive sentiments, indicating overall customer satisfaction. However, the notable presence of negative reviews emphasizes areas needing improvement, such as service quality and pricing strategies. This distribution highlights opportunities for restaurants to further enhance their strengths and mitigate weaknesses identified in customer feedback.

#### **4.3 Classification Performance**

**Table 1: Random Forest Classification Report**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Random Forest | 92.46% | 0.93 | 0.92 | 0.92 |

The performance of the Random Forest Classifier was evaluated using key metrics such as accuracy, precision, recall, and F1-score. The classifier achieved an impressive overall accuracy of 92.46%, demonstrating its effectiveness in categorizing restaurant reviews. The precision value of 0.93 indicates that the model has a high rate of correctly identifying positive and negative reviews. However, the recall for negative reviews (55%) suggests a limitation in identifying all actual negative cases, likely due to class imbalance in the dataset. The F1-score of 0.92 reflects the harmonic mean of precision and recall, indicating the model's balanced and reliable performance for sentiment classification.

1. **CONCLUSION**

In conclusion, this study successfully applied a Random Forest Classifier to sentiment analysis of restaurant reviews, achieving high accuracy and demonstrating the model's effectiveness in identifying customer sentiments. The integration of textual and numerical features provided actionable insights into areas of satisfaction, such as food quality and service, and dissatisfaction, such as pricing and staff behavior. The sentiment distribution and word cloud visualizations revealed critical themes in customer experiences, offering clear avenues for service improvement. While the model showed strong performance, the lower recall for negative sentiments indicates opportunities for further optimization. Future research could explore advanced feature engineering, incorporate additional datasets, or implement hybrid models to enhance classification performance. By emphasizing scalability and interpretability, this approach can be effectively adopted in diverse operational contexts within the restaurant industry.

**REFERENCES:**

[1] Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135. [https://www.nowpublishers.com/article/Details/INR-011](%20https:/www.nowpublishers.com/article/Details/INR-011)

[2] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32 <https://link.springer.com/article/10.1023/A:1010933404324>

[3] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167 https://link.springer.com/book/10.1007/978-3-031-02145-9

[4] Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 168–177. https://doi.org/10.1145/1014052.1014073

[5] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 NAACL-HLT, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[6] Fan, R., Chang, K.-W., Hsieh, C.-J., Wang, X., & Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research, 9, 1871–1874.

[7] Yang, Z., Cai, Z., Li, Y., & Zeng, D. (2020). Applications of machine learning techniques for restaurant review sentiment analysis. Expert Systems with Applications, 158, 113541. https://doi.org/10.1016/j.eswa.2020.113541

[8] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 50, 723–762. [9] Fan, R., Chang, K.-W., Hsieh, C.-J., Wang, X., & Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research, 9, 1871–1874.

[9] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2), 267–307. https://doi.org/10.1162/COLI\_a\_00049

[10] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15–21. https://doi.org/10.1109/MIS.2013.30

[11] Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-based sentiment analysis of teachers’ evaluation. Applied Computational Intelligence and Soft Computing, 2016, 1–8. <https://doi.org/10.1155/2016/2385429>

[12] Saleh, K., Gaber, T., Hossny, M., & Nahavandi, S. (2015). Real-time sentiment analysis. Procedia Computer Science, 73, 254–263. https://doi.org/10.1016/j.procs.2015.12.002

[13] Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Transactions on Information Systems (TOIS), 26(3), 1–34.

[14] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning word vectors for sentiment analysis. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 142–150.

[15] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 1631–1642.

[16] Moh, T.-S., & Zhang, C. (2001). Improving random forest for classification. Proceedings of the 8th International Conference on Machine Learning and Applications, 543–548.

[17] Elrahman, S. M. A., & Abraham, A. (2013). A review of class imbalance problem. Journal of Network and Innovative Computing, 1(2), 332–340.

[18] Jindal, N., & Liu, B. (2008). Opinion spam and analysis. Proceedings of the International Conference on Web Search and Data Mining (WSDM), 219–230. https://doi.org/10.1145/1341531.1341560

[19] Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). SentiBench – A benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Science, 5(1), 1–29. <https://doi.org/10.1140/epjds/s13688-016-0085-1>

[20] Ho, T. K. (1995). Random decision forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, 278–282.

[21] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC), 1320–1326.

[22] Zahoor, Bawany & Hamid. Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning

[23] Kalarani, Deepika, & Tamilmani. Enhancing Review System of Restaurant Using Deep Learning Approach for Sentiment Analysis

[24] Abdullah, Waheed, & Hossain. Sentiment Analysis of Restaurant Reviews Using Machine Learning

[25] Sharif, Hoque, & Hossain. Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes

[26] Hossain, Sharif, &  Hoque. SentiLSTM: A Deep Learning Approach for Sentiment Analysis of Restaurant Reviews

[27] Restaurant Reviews. <https://www.kaggle.com/datasets/gorororororo23/european-restaurant-reviews>