

COMPARATIVE STUDY OF SENTIMENT CLASSIFICATION MODELS ON RESTAURANTS REVIEWS DATASET

Swati Singh¹, Priya Pathak², Pooja Pandey³

¹Assistant Professor, Data Science, Thakur College of Science and Commerce, Mumbai, Maharashtra, India.

^{2,3}PG Student, Data Science, Thakur College of Science and Commerce, Mumbai, Maharashtra, India.

ABSTRACT

Customer reviews provide valuable insights into service quality, helping businesses improve their offerings. This study applies Aspect-Based Sentiment Analysis (ABSA) to restaurant reviews, identifying sentiments towards key aspects such as food, service, and ambience.

A rule-based approach is used for aspect extraction, where predefined keyword lists categorize review content. The dataset undergoes preprocessing, including text cleaning and stopword removal, to enhance accuracy. For sentiment classification, six machine learning models—Logistic Regression, Random Forest, SVM, Naïve Bayes, KNN, and xgboost—are trained and tested. Among them, SVM achieved the highest accuracy of 92.42% in classifying sentiments as positive or negative.

The findings offer structured insights into customer experiences, enabling restaurant owners to identify strengths and areas for improvement. This approach ensures efficient and interpretable sentiment analysis, providing a scalable framework that can be adapted to other domains detailed customer feedback analysis.

Keywords: Yelp Reviews, Machine Learning, Natural Language Processing (NLP), TF-IDF Vectorization, Restaurant Feedback.

1. INTRODUCTION

Customer reviews are essential for understanding the strengths and weaknesses of a restaurant. They provide direct feedback from diners, highlighting aspects such as food quality, service, and ambience. However, manually analyzing a large number of reviews is time-consuming and inefficient, making it difficult for restaurant owners to extract meaningful insights. To address this challenge, this project focuses on Aspect-Based Sentiment Analysis (ABSA) to systematically analyze customer feedback. By identifying sentiments related to specific aspects, restaurant owners can gain a clear understanding of what customers appreciate and what needs improvement. This structured approach helps in making informed decisions to enhance the overall dining experience. By automating the analysis of restaurant reviews, this study provides a practical and efficient solution for understanding customer opinions. The insights obtained can assist in making targeted improvements, ensuring better service and higher customer satisfaction.

2. LITERATURE REVIEW

1. The paper "Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques" by Akshay Krishna and co-authors explores how machine learning algorithms can effectively analyze customer sentiments in restaurant reviews. Sentiment analysis, a key technique in opinion mining, helps categorize opinions from text data into positive, negative, or neutral sentiments. The paper addresses the increasing need for businesses to efficiently interpret vast amounts of user-generated content to enhance services. The authors used a restaurant review dataset, which was pre-processed by removing punctuation, filtering stop words, and applying the Porter Stemmer algorithm to reduce words to their root forms. They implemented a bag-of-words model to convert the reviews into a binary vector format, enabling feature extraction for training the machine learning models. Various algorithms, including Naïve Bayes, Support Vector Machine (SVM), Decision Tree, and Random Forest, were applied, with the dataset split into training and testing sets. The SVM model achieved the highest accuracy of 94.56%, outperforming the other classifiers. Performance was evaluated through confusion matrices, False Acceptance Rates (FAR), False Rejection Rates (FRR), and accuracy metrics. The paper concludes that SVM is the most suitable model for sentiment analysis in this context, demonstrating the practical value of these techniques in helping businesses understand customer feedback and make informed decisions.

2. The paper "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes" by Rachmawan Adi Laksono, Kelly Rossa Sungkono, Cahyaningtyas Sekar Wahyuni, and Riyanarto Sarno analyzes customer reviews from TripAdvisor. The authors collected 337 reviews of Surabaya restaurants using a tool called WebHarvy. They cleaned the data by removing unnecessary words and punctuation. Then, they applied the Naïve Bayes algorithm and compared it with a simpler tool, TextBlob, to classify the reviews as positive or negative. Naïve Bayes, implemented in the WEKA software, achieved an accuracy of 72.06%, while TextBlob achieved 69.12%. The results were evaluated

using accuracy, precision, and recall, and Naïve Bayes performed better overall. The authors suggest that more data or other methods could improve future results.

3. The paper "Sentiment Analysis of Restaurant Reviews" by R. Rajasekaran, Uma Kanumuri, M. Siddhardha Kumar, Somula Ramasubbareddy, and S. Ashok focuses on sentiment analysis of restaurant reviews to classify customer opinions as positive or negative. The authors used a dataset of reviews and implemented the Naïve Bayes algorithm, achieving an accuracy of 72.2%. The study highlights how sentiment analysis helps businesses understand customer feedback and improve services. The authors also compared Naïve Bayes with Support Vector Machines (SVM), noting that SVM performs better geometrically, but Naïve Bayes is simpler and effective for tasks like spam detection. The paper concludes by emphasizing the importance of sentiment analysis in decision-making and recommends future work on addressing challenges like sarcasm, slang, and multimodal analysis.

4. The paper "Understanding Customer Sentiment: Lexical Analysis of Restaurant Reviews" by Jinat Ara, Md. Toufique Hasan, Abdullah Al Omar, and Hanif Bhuiyan presents a method to analyze restaurant reviews using Natural Language Processing (NLP). The dataset for the analysis was collected through a captive portal on a restaurant's web page, where customers shared opinions on food quality, service, and the environment. The authors employed the SentiStrength classifier to calculate the sentiment polarity (positive, negative, or neutral) of the reviews. The results were further refined using the Standard Deviation (SD) method to classify the strength of opinions. From a dataset of 700 opinions, the approach achieved an accuracy of 85.71% in predicting customer sentiment. The paper concludes that automated sentiment analysis provides valuable insights for restaurants to improve customer satisfaction.

5. The paper "The Cultural Impact on Social Commerce: A Sentiment Analysis on Yelp Ethnic Restaurant Reviews", by Makoto Nakayama and Yun Wan, analyzes how cultural differences affect customer reviews on Yelp for Japanese restaurants. Using 76,704 English-language and 56,159 Japanese-language reviews from Yelp, the authors examine the cultural impact on sentiments expressed in reviews regarding food quality, service, ambiance, and price fairness. Bilingual text mining software was used to process and analyze the reviews. The results reveal that Japanese reviewers focus more on food quality and price fairness, while Western reviewers emphasize service and ambiance. The study concludes that businesses must consider cultural differences when interpreting online reviews, as they directly influence customer perceptions. The findings offer practical insights for businesses looking to improve their services in multicultural social commerce environments, helping them better understand how cultural context shapes customer feedback.

6. The research paper "Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: A Systematic Review" examines how sentiment analysis can help food delivery services (FDS) improve customer satisfaction. With the rise of online food delivery, companies rely on customer feedback to assess service quality. Traditional approaches, such as lexicon-based methods and machine learning (ML) models like Naïve Bayes and SVM, have been widely used but often lack accuracy. Deep learning (DL) models, including Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN), provide higher accuracy but suffer from a lack of interpretability. To address this challenge, the paper explores Explainable Artificial Intelligence (XAI) techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP). LIME provides instance-based interpretations, while SHAP assigns contribution scores to input features, offering better transparency. The study reviews 97 research papers published between 2001 and 2022 and finds that 77% of sentiment analysis models in FDS lack interpretability, making it difficult for businesses to trust AI-driven decisions. The study suggests that combining DL models with XAI can improve both accuracy and explainability. Additionally, topic modeling can categorize negative reviews into key complaint areas such as delivery time, customer service, food quality, and cost, allowing businesses to take targeted actions. The paper concludes that future research should focus on hybrid DL models, domain adaptation, and integrating XAI for better interpretability. By adopting explainable deep learning approaches, FDS companies can enhance customer satisfaction and build trust in AI-driven sentiment analysis. This study provides a valuable roadmap for improving AI-powered customer feedback analysis in the food delivery industry.

7. The research paper "Sentiment Polarity Analysis of Bangla Food Reviews Using Machine and Deep Learning Algorithms" explores sentiment analysis in Bangla food reviews to assess food quality. The study collects 1,484 reviews from online food delivery platforms such as FoodPanda and HungryNaki to determine whether reviews are positive or negative. The dataset undergoes preprocessing steps such as tokenization, stopword removal, and vectorization using CountVectorizer, TF-IDF, and N-gram techniques. Various machine learning (ML) and deep learning (DL) models are applied, including Logistic Regression, Naïve Bayes, Random Forest, SVM, Decision Tree, and LSTM. Among them, Logistic Regression achieved the highest accuracy of 90.91%. The study highlights the importance of sentiment analysis for improving food service quality and providing insights into customer satisfaction. It suggests that restaurant owners

can benefit from analyzing online reviews to understand customer preferences and issues. The authors propose future enhancements such as expanding the dataset, implementing BERT or Transformer models, and incorporating multi-class classification to include neutral sentiments. This research significantly contributes to Bangla sentiment analysis by providing a labeled dataset and evaluating multiple models, offering a foundation for further improvements in food review sentiment classification.

8. The research paper "Deep Learning Approach for Sentimental Analysis of Hotel Reviews on Bengali Text" explores sentiment analysis of hotel reviews written in Bengali to help hotel management understand customer feedback. Since most sentiment analysis studies focus on English, this paper aims to develop a model that can classify Bengali hotel reviews as positive or negative. The authors collected 1,250 reviews from social media platforms, hotel websites, and surveys, translating reviews when necessary. The dataset underwent preprocessing, including stopword removal, tokenization, and stemming. Various feature extraction techniques were applied, such as TF-IDF, Count Vectorization, Word2Vec, and GloVe embeddings. The study employed both machine learning models (Logistic Regression, SVM, Naïve Bayes, Random Forest) and deep learning models (RNN, LSTM, GRU) for classification. Among them, LSTM achieved the highest accuracy of 76.64%, outperforming other models in detecting sentiment. The study highlights the importance of sentiment analysis for improving hotel services and customer experience. Future work aims to expand the dataset, include multi-class classification (e.g., very happy, normal, very bad), and develop a lightweight sentiment classification model for low-resource environments. This research is one of the first studies focusing on Bengali hotel reviews, making a significant contribution to Bangla Natural Language Processing (NLP) and sentiment analysis.

9. The research paper "Analyzing Customer Reviews on Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence (XAI)" explores sentiment analysis of customer feedback in the food delivery service (FDS) industry. With the rise of online food delivery platforms like UberEATS, Menulog, and Deliveroo, analyzing customer reviews has become crucial for understanding service quality and improving customer satisfaction. The study aims to develop an AI-based sentiment analysis framework using deep learning (DL) models and Explainable Artificial Intelligence (XAI) techniques to enhance model transparency. The research follows a threefold approach: (1) Sentiment classification using Long Short-Term Memory (LSTM), Bi-LSTM, and a hybrid Bi-GRU-LSTM-CNN model, achieving a high accuracy of 96.33%. (2) Applying XAI techniques (SHAP and LIME) to interpret DL model predictions and build trust in AI-driven insights. (3) Using Latent Dirichlet Allocation (LDA) for topic modeling to categorize customer complaints into key areas like delivery time, food quality, and customer service. The findings suggest that DL models outperform traditional ML techniques but require XAI for better interpretability. Future work includes expanding datasets, improving interpretability, and integrating advanced NLP models like BERT. This study provides a robust AI framework for sentiment analysis in FDS, bridging the gap between accuracy and explainability in deep learning.

3. METHODOLOGY

The methodology for aspect-based sentiment analysis consists of multiple phases: data preprocessing, aspect segmentation, feature extraction, and classification.

3.1 Data Preprocessing

The yelp restaurants review dataset used in this study consists of customer reviews with sentiment labels. The preprocessing pipeline includes:

Handling Missing Values: Reviews with null or missing values are removed to ensure clean data.

Text Cleaning: All text is converted to lowercase, and special characters, numbers, and punctuation marks are removed.

Tokenization: Each review is broken into individual words or tokens using the Natural Language Toolkit (NLTK).

Stop word Removal: Common words with little semantic meaning (e.g., "the," "and," "is") are removed using predefined stop word lists



Fig.1.Positive Review



fig.2.Negative Review

3.2 Feature Extraction

To convert textual data into numerical format suitable for machine learning, the study applies:

TF-IDF Vectorization: Term Frequency-Inverse Document Frequency (TF-IDF) is used to represent each review based on the importance of words in the corpus.

N-grams Representation: Bi-grams and tri-grams are used to capture context and improve sentiment classification accuracy.

Feature Selection: Features with low importance or high sparsity are removed to optimize model performance.

3.3 Rule-Based Sentiment Classification

Sentiment Lexicons: Predefined sentiment lexicons are used to assign sentiment scores to words and phrases.

Aspect-Sentiment Mapping: Rules are designed to match specific aspects with sentiment-laden words.

3.4 Machine Learning Classification

Logistic Regression: Uses a sigmoid function to classify reviews as positive, or negative. Works well with high-dimensional text data like TF-IDF. Fast, interpretable, but assumes linear relationships in data.

Random Forest: A collection of multiple decision trees, where each tree votes on sentiment. Reduces overfitting by averaging results across trees. Computationally expensive but handles complex data well.

Support Vector Machine (SVM): Finds the best decision boundary (hyperplane) to classify sentiments. Excels in high-dimensional text data, making it great for NLP tasks. Performed well in this study, indicating strong classification accuracy.

Naïve Bayes: Uses Bayes' Theorem to calculate the probability of a sentiment class. Fast and scalable for text-based classification. Works best when words are independent, but struggles with complex language.

K-Nearest Neighbors (KNN): Classifies reviews based on the sentiment of the closest neighbors. Non-parametric, meaning no fixed assumptions about data distribution. Slow for large datasets, as it must compare with all other samples.

XGBoosts: Boosting algorithm that optimizes performance by learning from errors. Highly efficient, reducing bias and variance. One of the best-performing models in this study, alongside SVM.

Hybrid Approach: Predictions from the machine learning models are combined with rule-based sentiment scores to improve accuracy.

Model Training and Evaluation: The dataset is split into training (80%) and testing (20%) subsets. The models were trained on labeled reviews (positive, negative, neutral). Performance was measured using accuracy, precision, recall, and F1-score. **SVM and XGBoost performed best**, showing strong classification results.

4. RESULTS & DISCUSSION

To assess the performance of different machine learning models for sentiment anal, we evaluated them using metrics like: Accuracy, Precision, and Recall, f1-score, support.

4.1 Model Performance Evaluation

The performance of six different machine learning models was evaluated for sentiment classification in restaurant reviews. The models were assessed based on **accuracy, precision, recall, and F1-score**. The following table summarizes the performance metrics:

Model	Accuracy	Precision (avg)	Recall (avg)	F1-Score (avg)
Logistic Regression	0.9189	0.92	0.92	0.92
Random Forest	0.8962	0.90	0.90	0.89
Support Vector Machine (SVM)	0.9242	0.92	0.92	0.92
Naïve Bayes	0.8850	0.89	0.88	0.88
K-Nearest Neighbors (KNN)	0.7923	0.80	0.79	0.77
XGBoost	0.8988	0.90	0.90	0.90

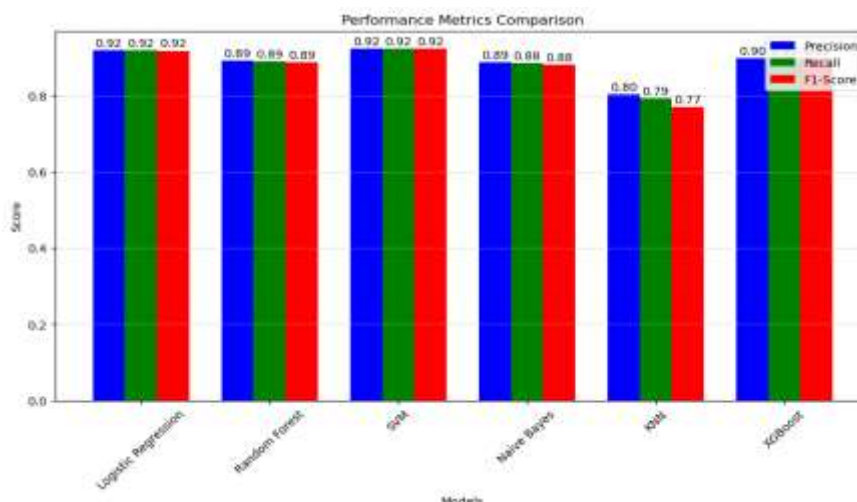
4.2 Best Performing Model

The performance evaluation of six machine learning models—Logistic Regression, Random Forest, Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbors (KNN), and XGBoost—revealed that SVM achieved the highest accuracy of 92.42%, making it the best-performing model for sentiment classification in restaurant reviews. This superior performance indicates that SVM effectively captures the complex decision boundaries between positive,

negative, and neutral sentiment classes, allowing for more precise sentiment predictions. SVM performed best due to several factors. First, it handles high-dimensional data efficiently, which is crucial for sentiment analysis as textual data is inherently sparse and high-dimensional. By leveraging kernel tricks, SVM transforms the feature space to enhance classification. Second, SVM is robust to overfitting, unlike tree-based models, which may struggle with generalization. This characteristic makes SVM a reliable choice for sentiment classification, even with a limited dataset. Third, SVM creates strong decision boundaries, ensuring a clear separation between different sentiment classes using optimized hyperplanes. While KNN demonstrated the lowest accuracy (79.23%), struggling with sentiment classification due to its reliance on distance-based similarity, other models like Naïve Bayes showed high recall (97%) for positive reviews but struggled with negative sentiment detection. Meanwhile, XGBoost and Random Forest performed well but were slightly outperformed by SVM and Logistic Regression. Additionally, the incorporation of Aspect-Based Sentiment Analysis (ABSA) significantly improved sentiment classification by analyzing sentiments associated with specific aspects of restaurant experiences. For instance, a review stating, "The service was slow but the food was delicious," would be categorized as negative for service but positive for food, offering more granular insights than overall sentiment classification. This detailed analysis allows restaurant owners to pinpoint strengths and weaknesses in their services, leading to more effective decision-making for customer satisfaction improvements. Despite achieving high accuracy with SVM, several areas can be explored for future improvements. The use of deep learning models such as LSTMs and BERT could enhance sentiment classification by capturing contextual information more effectively. Moreover, integrating sentiment intensity analysis could provide a sentiment score rather than simple categorical classification, enabling a more nuanced understanding of customer opinions. Further, expanding the model to analyze multilingual reviews would improve its applicability to diverse customer bases. Finally, deploying the model for real-time sentiment tracking could assist businesses in continuously monitoring and responding to customer feedback proactively.

4.3 Comparison of Results

SVM achieved the highest performance with an F1-score of 0.9446 (positive) and 0.8802 (negative), making it the best model. • Logistic Regression performed strongly, with F1-scores of 0.9408 (positive) and 0.8719 (negative), closely following SVM. • XGBoost showed competitive results, achieving F1-scores of 0.9263 (positive) and 0.8385 (negative). • Random Forest and Naïve Bayes performed moderately well, but had slightly lower accuracy than the top models. • KNN had the weakest performance, with an F1-score of 0.8618 (positive) and only 0.5822 (negative), struggling with negative sentiment. • Overall, SVM proved to be the best model, while KNN performed the worst.



This research explores sentiment analysis of Yelp restaurant reviews using machine learning models. The model was trained solely on preprocessed reviews with sentiment labels, without explicit aspect classification. However, during testing, predefined aspects (Food, Service, Ambience) were introduced, allowing the model to infer sentiment accordingly. Among the evaluated models, SVM and Logistic Regression outperformed others with an F1-score of 0.94, followed by XGBoost (0.92). The study highlights the impact of effective preprocessing (lemmatization, TF-IDF) on model performance and demonstrates that aspect-based sentiment can be accurately predicted even when not explicitly trained, offering valuable insights for businesses.

4.4 Aspect-Based Sentiment Analysis (ABSA) Impact

By implementing Aspect-Based Sentiment Analysis (ABSA), sentiment classification was further improved as it enabled more granular analysis of customer reviews. For example:

"The service was slow but the food was delicious."

Aspect 1: Service → Negative

Aspect 2: Food → Positive.

```
* **Aspect-Based Sentiment Analysis Results:**
* **Comment:** 'The pizza was amazing,' → **Aspects:** ['Food'] → **Sentiment:** positive
* **Comment:** 'the service was extremely slow' → **Aspects:** ['Service'] → **Sentiment:** negative
* **Comment:** 'The waiter was rude,' → **Aspects:** ['Service'] → **Sentiment:** negative
* **Comment:** 'the ambience was nice' → **Aspects:** ['Ambience'] → **Sentiment:** positive
* **Comment:** 'The burger was cold' → **Aspects:** ['Food', 'Ambience'] → **Sentiment:** negative
* **Comment:** 'tasteless' → **Aspects:** ['Food'] → **Sentiment:** negative
* **Comment:** 'the staff was very friendly,' → **Aspects:** ['Service'] → **Sentiment:** positive
* **Comment:** 'the restaurant had a great atmosphere' → **Aspects:** ['Ambience'] → **Sentiment:** positive
* **Comment:** 'The pasta was too salty,' → **Aspects:** ['Food'] → **Sentiment:** negative
* **Comment:** 'the drinks were overpriced' → **Aspects:** ['General'] → **Sentiment:** negative
* **Comment:** 'I really liked the music' → **Aspects:** ['Ambience'] → **Sentiment:** positive
* **Comment:** 'the lighting was beautiful' → **Aspects:** ['Ambience'] → **Sentiment:** positive
* **Comment:** 'The sushi was fresh' → **Aspects:** ['Food', 'Ambience'] → **Sentiment:** positive
* **Comment:** 'delicious,' → **Aspects:** ['Food'] → **Sentiment:** positive
* **Comment:** 'the waiter took too long to bring our drinks' → **Aspects:** ['Service'] → **Sentiment:** negative
* **Comment:** 'The place was beautifully decorated' → **Aspects:** ['Ambience'] → **Sentiment:** positive
```

5. CONCLUSIONS

This study analyzed 48,000 Yelp restaurant reviews using machine learning for sentiment classification. Among the models tested, Support Vector Machine (SVM) performed best, achieving the highest F1-score (0.9446 for positive, 0.8802 for negative sentiments). Logistic Regression and XGBoost also showed strong performance, while Random Forest and Naïve Bayes were moderate, and K-Nearest Neighbors (KNN) performed the weakest. The results highlight the importance of choosing the right model for sentiment analysis, with SVM proving most effective. Future research can explore deep learning and aspect-based sentiment analysis for deeper insights into customer opinions on food, service, and ambience.

6. REFERENCES

- [1] Krishna, A., Akhilesh, V., Aich, A., Hegde, C. (2019). Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques. In: Sridhar, V., Padma, M., Rao, K. (eds) Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol 545.
- [2] R. A. Laksono, K. R. Sungkono, R. Sarno and C. S. Wahyuni, "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes," 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 2019, pp. 49-54
- [3] Rajasekaran, R., Kanumuri, U., Siddhardha Kumar, M., Ramasubbareddy, S., Ashok, S. (2019). Sentiment Analysis of Restaurant Reviews. In: Satapathy, S., Bhateja, V., Das, S. (eds) Smart Intelligent Computing and Applications . Smart Innovation, Systems and Technologies, vol 105.
- [4] Rajasekaran, R., Kanumuri, U., Siddhardha Kumar, M., Ramasubbareddy, S., Ashok, S. (2019). Sentiment Analysis of Restaurant Reviews. In: Satapathy, S., Bhateja, V., Das, S. (eds) Smart Intelligent Computing and Applications . Smart Innovation, Systems and Technologies, vol 105
- [5] J. Ara, M. T. Hasan, A. Al Omar and H. Bhuiyan, "Understanding Customer Sentiment: Lexical Analysis of Restaurant Reviews," 2020 IEEE Region 10 Symposium (TENSYP), Dhaka, Bangladesh, 2020,
- [6] Nakayama, M., & Wan, Y. (2018). The cultural impact on social commerce: a sentiment analysis on Yelp ethnic restaurant reviews. Information & Management.
- [7] Adak A, Pradhan B, Shukla N. Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. Foods. 2022 May 21;11(10):1500.
- [8] Amin A, Sarkar A, Islam MM, Miazee AA, Islam MR, Hoque MM. Sentiment polarity analysis of bangla food reviews using machine and deep learning algorithms. In 2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) 2024 Apr 25 (pp. 1-6). IEEE.
- [9] Bonny JJ, Haque NJ, Ulla MR, Kanungoe P, Ome ZH, Junaid MI. Deep learning approach for sentimental analysis of hotel review on bengali text. In 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) 2022 Apr 21 (pp. 1-7). IEEE.
- [10] Adak A. Analyzing Customer Reviews on Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence (XAI). University of Technology Sydney (Australia); 2022.