

Hotel Sentiments Explored: A Deep Dive into Customer Review

Prof. Regina Fernandes¹, Siddhi Pusalkar², Aishwarya Gadhave³, Rohan Zanje⁴ Atharva Sawant⁵
^{1,2,3,4,5}Student, Computer Engineering Department, Terna Engineering College, Nerul, India

Abstract- *Sentiment Analysis is a machine learning technique designed to interpret human emotions. By enabling machines to comprehend and extract insights from emotions, it becomes a valuable resource for business growth and development. Hotel reviews gathered from guests can be categorized as positive, negative, or neutral, allowing for sentiment analysis. This concise analysis of reviews is crucial for maintaining quality control in hotel services. This project adopts an advanced approach to extracting insights from hotel reviews by leveraging various machine learning algorithms along with BERT (Bidirectional Encoder Representations from Transformers), a cutting-edge natural language processing model. In today's digital era, where online reviews significantly impact consumer decisions, this project aims to innovate how the hospitality industry perceives and responds to customer sentiments, ultimately enhancing guest experiences and overall satisfaction.*

Keyword- Sentiment Analysis, BERT (Bidirectional Encoder Representation from Transformer)

1. INTRODUCTION

The significance of online reviews cannot be overstated; they are crucial for a hotel's presence on online platforms, serving as a valuable endorsement from satisfied guests. This, in turn, translates into increased business and revenue. Effective management of a hotel's brand on online portals instills confidence in potential customers, encouraging them to choose the hotel without hesitation. Categorizing reviews to extract insights has become an integral aspect of the hotel industry. Reviews narrate the customer's perception of the services provided by the hotel. Positive reviews can be leveraged to highlight the hotel's commendable efforts, while negative feedback should be acknowledged and addressed promptly. Sentiment analysis is indispensable for automating the process of discerning whether a review expresses a positive, negative, or neutral opinion about the hotel and its services. By employing sentiment analysis,

hotels can streamline the task of evaluating customer data, including reviews, ratings, and social media comments.

Various rule-based methods exist for sentiment analysis utilizing Natural Language Processing (NLP) techniques such as parsing, stemming, and tokenization, in conjunction with manually crafted rules. Initially, two lists comprising positive and negative word parameters are defined. For instance, positive words like "decent," "excellent," and "lovely," and negative words like "poor," "terrible," and "horrible." Subsequently, a rule-based system processes these predefined word lists, tallying the occurrences of positive, negative, and neutral sentiments within the review. If the count of negative words exceeds that of positive words, the system classifies the sentiment as negative, and vice versa. Based on the imbalance between positive and negative words, sentiments are classified as positive or negative, aiding hotels in gauging customer satisfaction levels and identifying areas for enhancement.

2. RESEARCH METHODOLOGY

2.1 Surveying Existing System

After consulting various research papers, it became evident that Sentiment Analysis holds significant potential across various applications and industries, including enhancing customer experiences, revitalizing brand value, and monitoring social media comments. Through this research, numerous features of Sentiment Analysis have been identified, offering deep insights into business channels. Our investigation revealed a substantial growth in research focused on opinion mining, text classification, and sentiment analysis in recent years. According to available data, nearly 11 thousand papers on relevant topics have been published, with over 95% emerging after 2004, marking Sentiment Analysis as one of the fastest-growing areas of interest for researchers. Notably, a majority of these papers

concentrate on sentiment analysis research articles, indicating widespread attention from the general public as well.

Interestingly, historical data shows that studies measuring public opinions, particularly those conducted during and after World War II, were primarily motivated by legislative concerns. However, the surge in interest in sentiment analysis, particularly related to user sentiments expressed through online reviews and comments, began in the early to mid-2000s. Since then, sentiment analysis has expanded to encompass various areas, including predicting brand value of products and analyzing reactions to events such as terrorist attacks. Moreover, ongoing research in sentimental analysis and natural language processing has addressed numerous challenges, including sarcasm detection and multilingual support, further enhancing the utility of sentiment analysis in diverse contexts.

2.2 Objective

- **Customer Insights:** Gain insights into customer opinions and feelings towards specific hotels. This information can help hotels understand their strengths and weaknesses from a customer perspective.
 - **Service Improvement:** Identify areas where hotels are performing well and areas that need improvement. Positive sentiment can highlight successful services and amenities, while negative sentiment can point out problems that need to be addressed.
- **Reputation Management:** Monitor and manage online reputation by tracking sentiment trends across different platforms. Address negative sentiments promptly to mitigate their impact on the hotel's image.
- **Competitor Analysis:** Compare sentiment scores across different hotels or competitors to benchmark performance and identify areas of differentiation.
- **Customer Satisfaction:** Assess the overall satisfaction level of customers with their hotel experience and identify common pain points.

3. PROPOSED SYSTEM

The proposed system for hotel sentiment analysis is a comprehensive solution designed to extract actionable insights from guest reviews to improve service quality and enhance guest satisfaction. Initially, the system collects data from various online platforms, including review websites and social media channels, to gather a diverse range of guest feedback. Once collected, the raw text data undergoes rigorous preprocessing, which

involves removing special characters, numbers, and irrelevant symbols, as well as tokenization, stopword removal, and stemming or lemmatization to ensure consistency and accuracy in analysis.

Following preprocessing, the system employs advanced feature extraction techniques, such as BERT (Bidirectional Encoder Representations from Transformers), to convert the preprocessed text into numerical features. This step enables the system to capture the nuanced meanings and context within guest reviews, allowing for more accurate sentiment analysis. For model training, a suitable machine learning algorithm is selected based on the nature of the data and the desired outcomes. Options range from deep learning models like LSTM (Long Short-Term Memory) or BERT to traditional classifiers like Naive Bayes or Support Vector Machines.

Once the model is trained using labeled data containing positive, negative, and neutral reviews, it undergoes evaluation to assess its performance using metrics such as accuracy, precision, recall, and F1-score. The sentiment analysis phase involves utilizing the trained model to predict sentiments of new hotel reviews and categorizing them into positive, negative, or neutral sentiments. These sentiment predictions are then leveraged to identify trends and patterns in guest feedback, enabling hotel management to pinpoint areas for improvement in service delivery, staff performance, and overall guest experiences. Visualizations such as charts and graphs are generated to illustrate sentiment trends over time, and comprehensive reports are generated to summarize sentiment distribution and key findings. Through the integration of this sentiment analysis system into hotel management software or customer feedback platforms, hoteliers gain access to real-time insights and feedback, empowering them to make data-driven decisions and enhance guest satisfaction.

4. LITERATURE REVIEW

The literature survey reveals related works in the field of sentiment analysis of hotel reviews. The first work focuses on the classification of hotel reviews from the goibibo.com website. The author proposed a model that classifies each review into one of the categories: Very good, Good, Average, Bad, Very Bad based on the sentiment value. The model was trained using 80% of the data and tested with the remaining 20%. The sentiment analysis was carried out by labeling

reviews with positive sentiments (e.g. happy, amazing, tasty, nice, pretty) and negative sentiments (e.g. bad, disgusting, sad, disappointed). The aim of this analysis is to provide suitable recommendations to customers and successful decision-making to business owners using sentiment-based results.

The second work conducted a sentimental analysis of hotel reviews from TripAdvisor. The proposed classification model takes hotel reviews as input and classifies each review into one of the categories based on the sentiment value. The model was tested with large data retrieved from web pages. The sentiment analysis was carried out by labeling reviews with positive sentiments (e.g. happy, amazing, tasty, nice, pretty) and negative sentiments (e.g. bad, disgusting, sad, disappointed). The goal of this analysis is to provide suitable recommendations to customers and successful decision-making to business owners using sentiment-based results.

The third work conducted by Sanjay, P., Nagarjuna Reddy, N., R.V. Ravi, Chand, K. Pujitha, and Anjali Mathur in 2019 focused on the implementation of machine learning algorithms for sentiment analysis in the context of hotel ratings.

The researchers used data collected from various websites such as Trip Advisor and Trivago to train and test their algorithms. Specifically, they utilized the Naïve Bayes classifier to categorize sentiments into two broad categories: positive and negative. The dataset consisted of sentimental statements extracted from various reviews. This research contributes to the growing body of literature on the application of machine learning techniques for sentiment analysis, particularly in the hospitality industry. By accurately classifying sentiments, businesses can gain valuable insights into customer experiences and make informed decisions to improve their services.

5. METHODOLOGY

Sentiment analysis techniques can be broadly categorized into three major categories: Statistical methods, Knowledge-based methods, and Hybrid techniques.

Statistical Methods: Also known as evolutionary approaches, statistical methods aim to find the mutual relationship between words sharing the same context. These methods leverage mathematical representations

of the text corpus. One simple approach is to sort a list of, say, 500 words that occur most frequently in positive texts while excluding negative ones, and vice versa. Models are then trained to determine whether there are more positive or negative words, without relying on linguistic insights.

Knowledge-based Methods: The goal of knowledge-based methods is to extract knowledge by classifying text based on explicitly present word categories, such as "awesome," "sad," "happy," "unfortunate," and "poor." These methods also extract knowledge from less obvious words, such as "sympathy," which is associated with particular emotions.

Hybrid Techniques: Hybrid techniques combine both statistical learning and knowledge-based methods to calculate polarity scores. This approach aims to achieve high accuracy and stability by leveraging the strengths of both methods. In our quest for accurate sentiment analysis on hotel reviews, we harnessed the power of BERT, a state-of-the-art pre-trained language model, coupled with three distinct algorithms: Support Vector Machine (SVM), Random Forest, and Naive Bayes. This multifaceted approach allowed us to explore diverse strategies and methodologies, optimizing our model for nuanced sentiment understanding in the context of hotel reviews. The integration of BERT, renowned for its contextualized word embeddings, provided our model with a deep understanding of the semantic intricacies present in natural language.

BERT (Bidirectional Encoder Representations from Transformers):

BERT is a natural language processing (NLP) model based on the Transformer architecture. It is designed to pre-train on large amounts of unlabeled text data bidirectionally, allowing it to understand the contextual relationships between words in a sentence. BERT has achieved state-of-the-art performance in various NLP tasks and is known for its ability to generate deep contextualized word embeddings.

SVM (Support Vector Machine):

SVM is a supervised machine learning algorithm used for classification and regression tasks. In the context of classification, SVM works by finding the hyperplane that best separates data points belonging to different classes in a high-dimensional space. It aims to maximize the margin between different classes, and it can be used with various kernel functions to handle non-linear relationships in the data.

Naive Bayes:

Naive Bayes is a probabilistic machine learning algorithm based on Bayes' theorem. It is particularly popular for text classification tasks. Despite its simplicity, Naive Bayes assumes that features are conditionally independent given the class label, which allows for efficient training and prediction. It is often used for spam filtering, sentiment analysis, and document classification.

Random Forest:

Random Forest is an ensemble learning algorithm that combines multiple decision trees to make more accurate predictions. Each tree is constructed using a random subset of the training data, and the final prediction is determined by aggregating the predictions of individual trees (e.g., through a majority voting mechanism for classification tasks). Random Forest is known for its robustness, versatility, and ability to handle high-dimensional data.

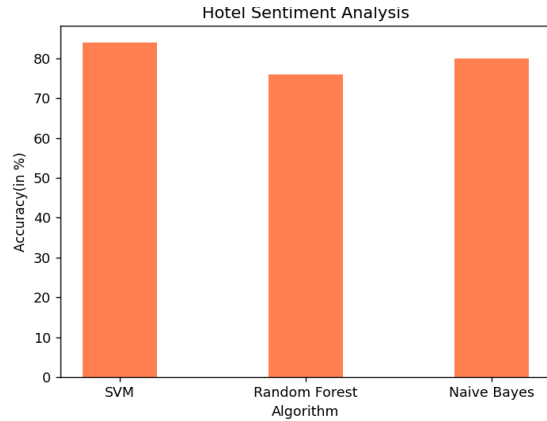
6. EVALUATION AND RESULT

The first algorithm we employed, Support Vector Machine (SVM), demonstrated exceptional synergy with BERT embeddings. The high accuracy of 84% suggests that SVM, with its capacity to discern complex decision boundaries, was adept at leveraging the rich contextual information provided by BERT. This combination yielded a robust sentiment analysis model capable of making precise predictions on the sentiment expressed in hotel reviews.

Our exploration extended to the utilization of Naive Bayes, a probabilistic algorithm known for its simplicity yet effectiveness. The BERT-Naive Bayes fusion achieved an accuracy of 80%, underscoring the adaptability of BERT embeddings with a more straightforward classification approach. This approach serves as a testament to the versatility of BERT in complementing various algorithmic strategies.

Furthermore, we integrated BERT with Random Forest, an ensemble learning algorithm. While the accuracy of 76% was slightly lower than the other combinations, it showcased the ability of BERT to synergize with diverse machine learning architectures, providing a holistic perspective on sentiment analysis.

ALGORITHM	ACCURACY
SVM	84%
NAÏVE BAYES	80%
RANDOM FOREST	76%



7. CONCLUSION

In the sentiment analysis conducted for hotel reviews, reviews were categorized into positive sentiments, which included words like "happy," "amazing," "tasty," "nice," and "pretty," and negative sentiments, which included words like "bad," "disgusting," "sad," and "disappointed." The objective of this analysis is to offer suitable recommendations to customers for selecting the best available option and to aid business owners in making informed decisions based on sentiment-based results and implications.

This analysis serves to determine the attitude of customers through online feedback on various aspects of hotel services, food quality, staff behavior, and ambiance. By leveraging sentiment analysis, businesses can gain valuable insights into customer perceptions and preferences, enabling them to tailor their offerings and services to better meet customer needs and expectations.

In conclusion, sentiment analysis of hotel reviews is a powerful tool for both customers and business owners. It allows customers to make informed decisions based on the sentiments expressed by others, while also providing valuable feedback to businesses for enhancing their offerings and improving overall customer satisfaction. By understanding and leveraging sentiment analysis, businesses can stay competitive in the hospitality industry and continuously strive for excellence in customer service.

8. ACKNOWLEDGMENTS

We are profoundly grateful to Prof. MINAL CHAUDHARI and Prof. REGINA FERNANDES for their expert guidance and continuous encouragement throughout to see that this project rights its target. At

last we must express our sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped us directly or indirectly during this course of work.

REFERENCE

- [1] Varshitha Ch, Rohit T, Jayanag Bayana, Sai Sriram T, Sumathi B Classification of Hotel Reviews using Sentiment Analysis Turkish Online Journal of Qualitative Inquiry (TOJQI) Volume 12, Issue 3, June 2021:1108- 1122
- [2] Sentimental Analysis of Hotel Reviews from TripAdvisor Vaibhav Singh¹, Aayushi Mahajan², Deepanshi Chaudhary³ International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 06 | June 2020
- [3] Sentiment Analysis for Hotel Reviews Vikram Elango and Govindrajan Narayanan
- [4] Learning Word Vectors for Sentiment Analysis by Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts
- [5] Hotel Review Sentiment Analysis using Machine Learning Mrs. V. Kalaivani, Shamini B, Srinidhi R, Tamil Selvi D International Research Journal of Modernization in Engineering Technology and Science (Peer-Reviewed, Open Access, Fully Refereed International Journal)Volume:05/Issue:03/March-2023