

# Analysis of Online Reviews for Customer Experience Insights

## Authors

Aishwarya Kumar Arvind, Gargi Girish Umrajkar, Raghul S, Vandit Gupta, Vinayaka Hosahalli Kotrappa

## Summary

The project focuses on extracting actionable insights from user-generated reviews on a major online platform to predict customer satisfaction and product quality through ratings. By analyzing reviews, ratings, and accompanying metadata, we aim to reveal patterns and sentiments that significantly impact customer experience. The primary objective is to develop a predictive model that can accurately forecast review ratings based on textual feedback. This analysis will enable businesses to enhance their services and assist consumers in informed decision-making. Early data exploration has underscored the dataset's potential for deep sentiment analysis and predictive modeling, confirming the project's feasibility and alignment with our goals.

## Proposed Plan

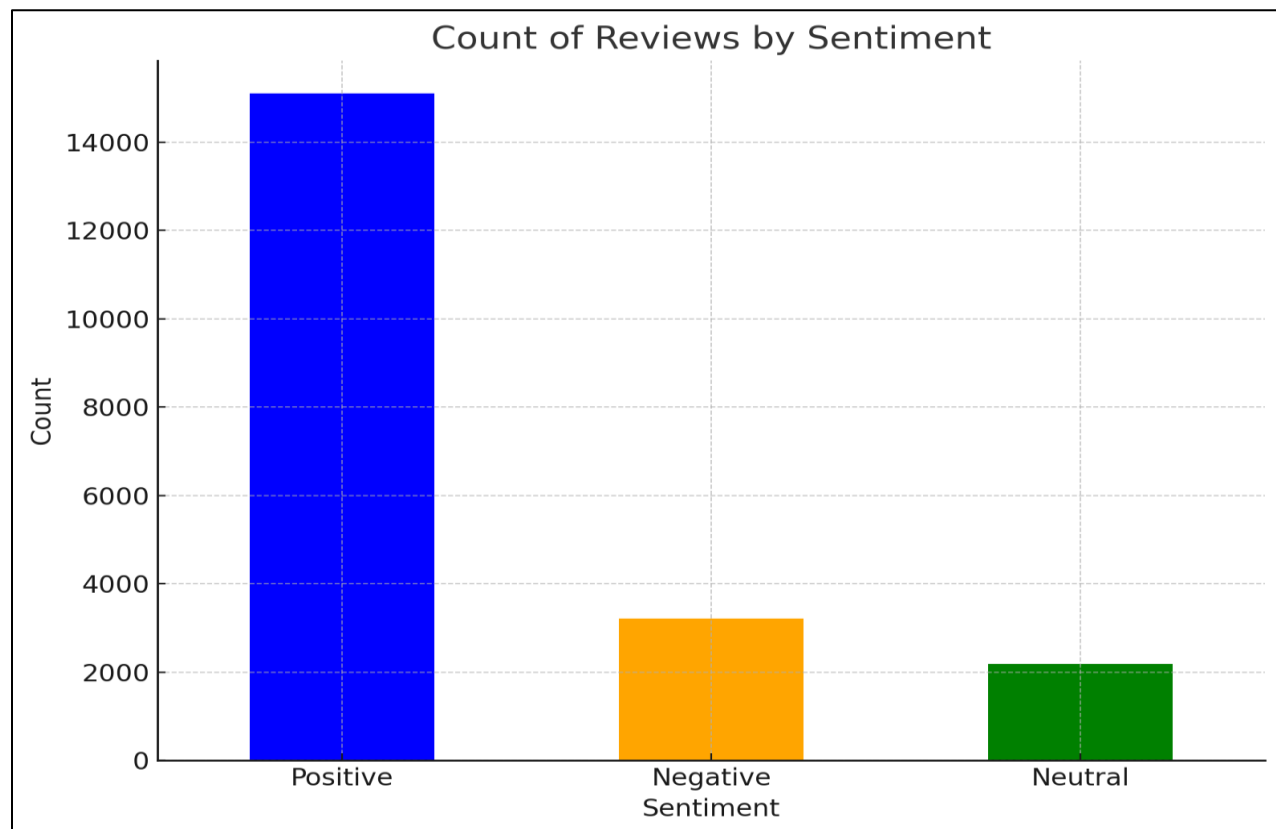
The approach will blend sentiment analysis, using VADER for initial sentiment categorization, with advanced text processing techniques (tokenization, lemmatization) via NLTK to prepare data for predictive modeling. The project's centerpiece is the development of a BiLSTM deep learning model, utilizing Keras, designed to predict the ratings of reviews by understanding the intricacies of textual sentiment. This approach addresses the significant challenge of correlating the nuanced sentiment of text data with numerical rating outcomes. Preliminary interactions with the dataset demonstrate its appropriateness for detailed sentiment analysis and the potential for accurate rating predictions.

## Preliminary Results

Initial exploration has successfully identified the dataset as a rich source for analysis, with comprehensive metadata supporting the project's predictive objectives. This preliminary phase has not only shown the data's compatibility with our analysis methods but also its promise for developing a reliable predictive model for review ratings.

To further demonstrate our capability to load and effectively explore the dataset, we present the head of the data in table form below, showcasing the initial rows of reviews and ratings, which affirms our proficiency in handling and understanding the dataset's structure and content. We have also included a bar chart visualizing the distribution of sentiments among the reviews. This chart succinctly illustrates the predominance of positive sentiments within the dataset and provides a clear, immediate visual understanding of the overall sentiment trends, reinforcing our proficiency in extracting and visualizing meaningful patterns from the data.

No	Review	Rating	Sentiment_Score	Sentiment
1	spacious rooms with excellent service	5	0.9632	Positive
2	average stay, unimpressive breakfast	3	0.6548	Neutral
3	poor cleanliness, would not recommend	1	0.2345	Negative
4	amazing amenities, friendly staff	5	0.9876	Positive
5	moderately priced, great location	4	0.7599	Positive



## References

1. Dataset: [https://zenodo.org/records/1219899#.YHwt1J\\_ivIU](https://zenodo.org/records/1219899#.YHwt1J_ivIU)
2. [https://www.researchgate.net/publication/350854936\\_Hotel\\_review\\_analysis\\_for\\_the\\_prediction\\_of\\_business\\_using\\_deep\\_learning\\_approach](https://www.researchgate.net/publication/350854936_Hotel_review_analysis_for_the_prediction_of_business_using_deep_learning_approach)
3. <https://medium.com/analytics-vidhya/predicting-the-ratings-of-reviews-of-a-hotel-using-machine-learning-bd756e6a9b9b>
4. <https://www.kaggle.com/code/jonathanoheix/sentiment-analysis-with-hotel-reviews>