Sentiment Analysis on Italian Hotel Reviews

Understanding Customer Feedback Through Data Science

Authors
Sharon Owino

Strathmore UNIVERSITY

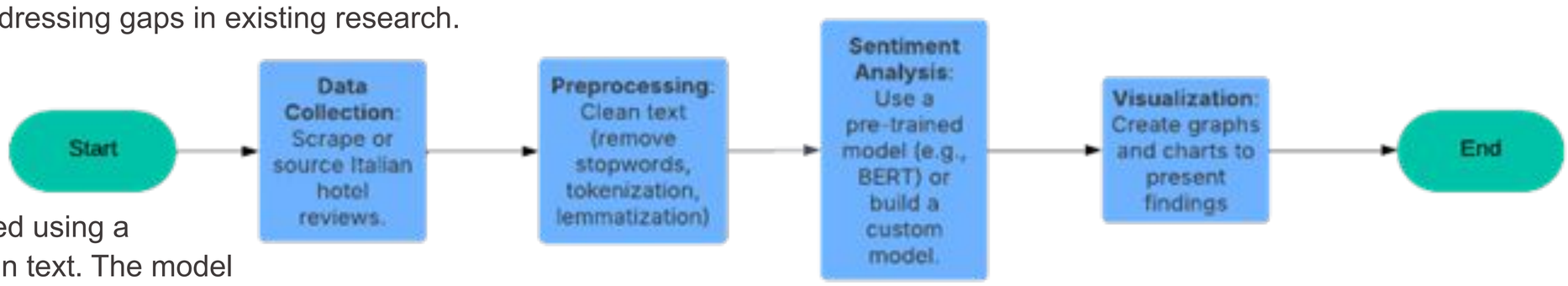
01 Introduction

Sentiment analysis, a natural language processing (NLP) technique, decodes emotions and opinions in reviews crucial in the hospitality industry and relies on customer satisfaction. This study explores predominant sentiments, service aspects influencing feedback, and variations across hotel types and seasons. By analyzing Italian reviews using advanced NLP like BERT, it offers language-specific insights and actionable recommendations, addressing gaps in existing research.

02 Objectives

- Analyze customer reviews to understand sentiment and its relationship with ratings.
- Identify factors that influence customer satisfaction based on reviews and ratings.
- Provide actionable insights to improve hotel services and customer experience.

03 Methodology

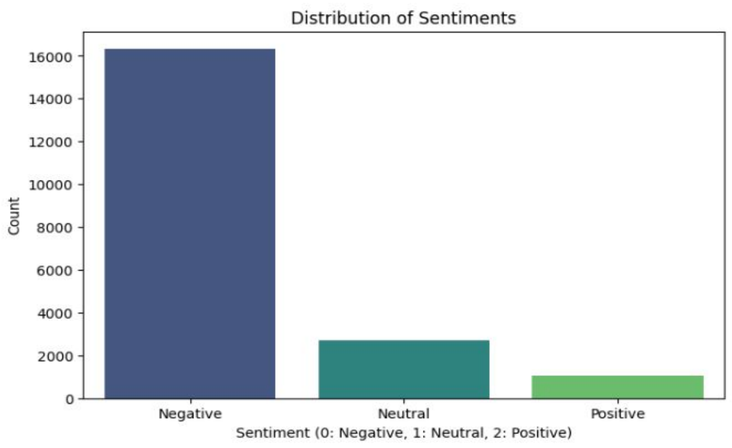


Sentiment analysis was conducted using a pre-trained BERT model for Italian text. The model was fine-tuned with AdamW optimizer, preprocessed data was converted into a TensorDataset, and validated and evaluated using a confusion matrix for performance metrics.

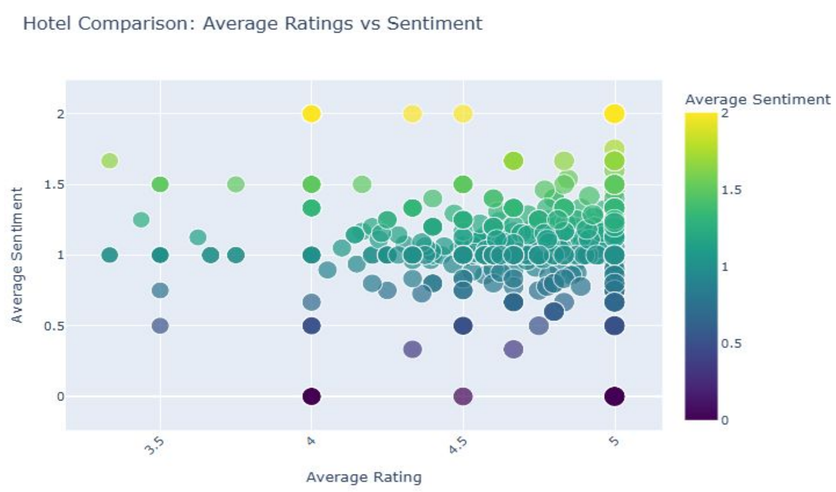
The figure above demonstrates the process for analyzing the dataset of Italian user comments from TripAdvisor, incorporating data collection, preparation, exploratory data analysis (EDA), and modeling.

04 Analysis

The BERT model effectively classifies sentiment in Italian hotel reviews reveals key factors influencing customer satisfaction, such as service quality and amenities.



The bar chart displays the distribution of sentiments across the dataset. It shows the count of Negative, Neutral, and Positive sentiments, with Positive being the most frequent.



The scatter plot explores the relationship between review length and sentiment scores. It helps determine if longer reviews are associated with stronger sentiments (positive or negative).

The line graph shows the average rating over time, revealing trends such as seasonal fluctuations or long-term improvements/declines in customer satisfaction.

05 Results / Findings

- The training loss curve shows effective model learning, with loss decreasing from 0.704 to 0.694.
- The confusion matrix highlights classification accuracy, while the ROC curve (AUC = 0.50) indicates room for improvement.
- Positive sentiments dominate, with longer reviews expressing stronger emotions.
- Heatmaps reveal hotel rating-sentiment correlations, and line graphs show seasonal and long-term rating trends. These findings provide actionable insights for enhancing hotel services based on customer feedback.

06 Conclusion

Key findings from using BERT for sentiment analysis to analyze Italian hotel reviews reveal that positive sentiments dominate, longer reviews express stronger emotions, and sentiment trends correlate with hotel ratings.

Recommendations include refining the model and addressing seasonal customer satisfaction trends for actionable insights.

Related Literature

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135. https://doi.org/10.1561/1500000011
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225.

GitHub Repositories

- Sentiment Analysis on Italian Hotel Reviews Poster. GitHub. Retrieved from https://github.com/sharonowino/Sentiment-Analysis-on-Italian-Hotel-Reviews
- Hugging Face. (n.d.). Transformers library. GitHub. Retrieved from https://github.com/huggingface/transformers
- Kaggle. (n.d.). Datasets. Retrieved from https://www.kaggle.com/ Articles
- Towards Data Science. (n.d.). Practical guides on BERT and sentiment analysis. Retrieved from https://towardsdatascience.com/