Predicting Hotel Rating based on the Hotel Reviews

Mojtaba Ghasemi

January 2024

1. Motivation and Problem

The impetus for this project stems from the pivotal role that online hotel reviews play in shaping the travel industry. In an era where digital platforms like Tripadvisor are increasingly influencing consumer choices, understanding the subtleties within customer reviews becomes crucial. The primary challenge lies in deciphering and quantifying the qualitative data from these reviews. This project seeks to address the gap in efficiently analyzing and interpreting large volumes of textual feedback to unveil patterns and insights that can significantly impact hotel ratings and customer satisfaction.

2. Potential Client

The insights gleaned from this project are poised to benefit a broad spectrum of clients in the hospitality and travel sector. Hoteliers and managers can leverage the findings to refine their services and offerings, aligning more closely with customer expectations. Online travel agencies and platforms, including Tripadvisor, can use the insights to enhance their recommendation algorithms and customer service strategies. Additionally, individual travelers, travel bloggers, and influencers can utilize this information to make more informed choices about accommodations, thereby enhancing their travel experiences.

3. Data Analysis:

This project will embark on a detailed analysis of 20,000 hotel reviews obtained from Tripadvisor. Utilizing cutting-edge Natural Language Processing (NLP) techniques, the project will focus on extracting critical features from these reviews, such as sentiment analysis, keyword extraction, and thematic categorization. This process will not only highlight the factors that most influence hotel ratings but also uncover underlying patterns and trends in customer preferences and expectations. The analysis will also explore correlations between various aspects of the reviews and the final ratings assigned by customers.

4. Solution and Approach:

To address the challenges of analyzing textual data, this project proposes a multi-faceted approach. Initially, the project will focus on feature engineering, identifying and extracting relevant data points from the reviews using NLP. Subsequently, a deep learning model will be developed and trained to predict hotel ratings based on these features. This model aims to be a sophisticated tool that can process and interpret the nuances of human language, thereby providing a more accurate and insightful understanding of customer feedback.

5. Deliverables:

Upon completion, the project will yield several key deliverables. A comprehensive analytical report will detail the methodology, findings, and implications of the study. A sophisticated deep learning model, capable of predicting hotel ratings from textual reviews, will be developed. This model will be accompanied by visual representations of the data insights, providing an accessible understanding of the findings. Additionally, Jupyter notebooks containing the complete code for data analysis and model development will be provided, enabling further exploration and application of the study's findings.

6. Source:

The dataset, consisting of 20,000 reviews from Tripadvisor, serves as the backbone of this project. Tripadvisor's status as a leading platform in the travel industry makes it an ideal source for diverse and comprehensive customer feedback. These reviews offer a rich tapestry of perspectives, experiences, and evaluations from travelers worldwide, providing a robust foundation for a detailed and insightful analysis.

https://zenodo.org/records/1219899#.YHwt1J_ivIU