

Bibliography.bib

Sorting sentiments of hotel reviews through machine learning

Darren Yap

Tan Jia He

Fu Jinghang

2024

contents

Abstract

Hotels depend on tourists to survive. They gather consumer opinions via customer reviews to improve their services. This project used sentiment analysis—the process of determining the emotional tone of text—to quantify consumers’ opinions on hotels, and predicted the overall sentiment of hotel reviews. International hotel reviews in English were split into individual words, assigning each word a score based on its relative intensity. The sentiment makeup of the review dataset was highlighted and the most frequent tokens were identified. Two machine learning models—a logistic regression model and a random forest classifier—were also constructed to predict the overall sentiment of a review. It was shown that the models were capable of predicting the overall sentiment of hotel reviews. This project highlights the possibility of using sentiment analysis models to create applications that allow customers to better understand the perceived quality of a hotel and potentially combat review fraud—the the dishonest practice of manipulating false reviews to boost a hotel’s ratings.

1 introduction

After the Singapore government relaxed travel restrictions due to COVID-19, there has been a recent increase in the number of tourists travelling in and out of Singapore. As such, hotels have seen a rise in the number of prospective tourists to be housed, and this may encourage an increase in the number of reviews hotels may receive.

Today, it is common to use social networks, messengers, and review websites to receive data from customer opinions. This is especially true for hotels, where previous occupants may evaluate the hotel on several factors through their reviews—be it cleanliness, facilities, location and convenience, etc. These come in two forms—a quantitative review (based on stars, diamonds, hearts, etc.) and a more qualitative review through text.

However, quantitative reviews do not always paint the full picture of customers’ opinions towards a certain hotel. Though it is certainly helpful to have a more objective rating system using numerical scores, eg. the Department of Tourism grading system in the Philippines, or the European Hotelstars Union system, these are given by customers subjectively and do not reflect the reasons for customers giving the rating. There is also evidence of manipulation of ratings by hotel management itself, where hotels may be compelled to forge positive or negative ratings to bias the overall rating. This made up 2.1 % of the 66 million reviews submitted to TripAdvisor in tripadvisor. Therefore, we propose using sentiment analysis to extract customers’ true feedback on hotels instead.