Text Classification of Reviews using Machine Learning Models, Review Summarization and Building Content-Based Hotel Recommender System

Himanshu Gupta, Master of Science in Computer Science
University of Dublin, Trinity College, 2020

Supervisor: Dr. Bahman Honari

Reviews are preliminary elements that travellers look at and analyse before booking any hotel. These reviews are of utmost importance for new customers as they can get insights about the hotels based on others' experiences. Moreover, it can benefit the hotel management staff so that they can upgrade the services and products based on the submitted feedback after thorough analysis. As per the Google trends, Europe was the most visited destination in 2018, approximately 710 million international tourists' arrivals occurred. Hence the number of travellers submitting the reviews was huge.

A widely faced challenge due to this huge chunk of reviews was that when a new traveller comes in, he would be reluctant to go through all the reviews which can lead to selecting an undesired hotel even though the ratings are considerably good. This outstanding challenge has motivated me to work on the European hotel dataset which would present the travellers with the most relevant reviews, which in turn would help them in choosing the best hotels in comparatively less time.

To achieve this, we resort to Review Classification method, which classifies the reviews into different classes (good and bad reviews) and further compare different machine learning models. In case the customer wants to analyse each hotel's review, it would be preferable to select the most relevant sentences from lengthy reviews. We will employ the Review Summarization technique to address this. Finally, customers may have some preferences like 'large room', 'couple-friendly' etc. based on which they would want to filter the hotels. This will be further accomplished by building a simple content-based recommender system. The results obtained in this study include – accurately classified reviews, precisely summarised huge reviews and recommended hotel list in order of relevance. Also, it was found that deep learning methods could classify reviews more accurately as compared to the other machine learning approaches.