문제 정의

코로나로 인하여 ‘호캉스’ 형태의 휴가가 증가하였습니다. 이로 인해 호텔이 단순 숙박시설이 아니라 하나의 여행지 역할을 하고 있습니다. 그만큼 소비자가 호텔을 결정하는 과정에서 이전보다 다양한 것을 고려할 것이라고 예상합니다. 이러한 상황에서 호텔이 소비자의 만족을 주는 요소가 무엇인지 살펴볼 수단이 필요하고 소비자들은 합리적 호텔 결정에 도움을 줄 수단이 필요하다고 생각했습니다. 일반적으로 호텔을 선정할 때 사람들은 호텔리뷰를 참고합니다. 호텔리뷰 분석을 통해 사람들이 호텔의 요소별 장단점을 확인할 수 있으며 고객은 호텔 예약 시, 이를 참고할 수 있습니다. 호텔 측의 경우 고객의 피드백으로 각 요소를 체크하며 품질 개선에 이를 활용할 수 있을 것입니다. 고객의 리뷰를 머신러닝 기술에 적용하여 호텔 평가를 하고자 합니다. 이 결과가 소비자와 호텔, 두 입장 모두에게 긍정적으로 작용할 수 있기를 기대합니다.

PROJECT INTRODUCTION:

Due to COVID-19, vacation in the form of "Staycation" has seen a steady increase. For this reason, hotels are not just accommodations, but serve as travel destinations. As such, consumers consider a variety of things in the process of deciding which hotel to stay in. In this situation, we thought that hotels needed a means to examine what factors satisfy consumers, and consumers needed a means to help rationalize accommodation decisions.

In general, when selecting a hotel, people refer to hotel reviews. Through hotel review analysis, people can check the strengths and weaknesses of each element of the hotel, and customers can refer to it when booking a hotel. On the hotel service side, each element or aspect can be checked according to customer feedback and used for quality improvement accordingly .

AIM: The aim of this project is to evaluate the hotel by applying machine learning technology in order to evaluate customer reviews. With the use of such technology, there can be improvement of quality and service on the hotel side whilst making the process of hotel selection on the part of the customer easier, thereby resulting in a positive effect on both consumers and hotel owners.

주제 선정

저희는 Kaggle에서 ‘Trip Advisor Hotel Reviews’라는 이름의 데이터를 이용합니다. 이 데이터는 Tripadvisior라는 사이트에서 크롤링한 20,000개의 리뷰의 데이터입니다. 우선, Data Set 속의 별점 수와 리뷰를 이용하여 해당 데이터를 긍정/부정 반응으로 분류했습니다. 즉 ‘Sentiment Analysis’를 이용해 소비자의 긍정/부정 태도에 대한 구분을 진행했습니다. 저희는 이제 LDA(Latent Dirichlet Allocation) 모델을 생성하여 소비자가 긍정적으로 생각하는 요소와 부정적으로 생각하는 요소를 구체적으로 살펴볼 것입니다. 이 결과를 참고하여, 호텔에게 긍정/부정적으로 작용하는 요소가 무엇인지 정리하는 것이 최종 목표입니다.

DATA COLLECTION/ STEPS TAKEN:

The dataset is collected from the ‘Trip Advisor Hotel Reviews’ on Kaggle. The data consists of 20,000 reviews crawled from a the travel website ‘Tripadvisor’.

First of all, the data was classified as positive or negative reactions depending on the number of stars and reviews in Data set. In other words, Sentimental Analysis is used to distinguish consumers' positive/negative attitudes.

We will now create a Latent Dirichlet Allocation (LDA) model to examine in detail the factors that consumers rate positively and negatively. Referring to these results, the final goal is to organize and distinguish which factors work positively or negatively for the hotel.