모델 1은 리뷰의 긍정, 부정을 판독하여 분류해주는 Text classification 모델이다. ‘Hotel\_Review.csv’ 데이터의 긍정적인 리뷰와 부정적인 리뷰 각각에 포함된 긍정어와 부정어의 비율을 구하고, 이를 input data로 하는 Random Forest를 구현하여 리뷰가 얼마나 긍정적인지, 혹은 부정적인지 판단하는 모델을 만들었다.

모델 2는 긍정, 부정 리뷰에 가장 언급된 키워드를 분석하여 소비자가 느끼는 기업의 제품이나 서비스의 강점과 약점을 구체적으로 보여주는 모델이다. 먼저 각 리뷰에 포함된 긍정, 부정어와 불용어들을 제외하였고 자연어 처리의 Part of speech tagging 알고리즘을 통해 각 단어의 품사를 labeling하고, 리뷰에서 중요한 명사와 형용사, 부사를 추출했다. 추출한 키워드들의 빈도수를 구함으로써 호텔의 강점과 약점을 찾아내었고 나아가 긍정, 부정, 그리고 공통적으로 나타난 단어들의 빈도 수를 시각화했다.