Smart Hotelier – Aspect Extraction and Sentimental Analysis on Hotel Reviews with BERT

Madhawa Dias

A Dissertation Submitted in Partial Fulfilment of the Requirement for
Bachelor of Science (Honours) Degree in Computer Science

**Department of Computing
Informatic Institute of Technology, Sri Lanka
in Collaboration with
University of Westminister, UK**

2020

Abstract

Text written in a natural language is difficult for a computer to interpret. Natural Language Processing(NLP) is a sub domain of Machine learning which allows a computer system to understand text written in human language.

User reviews are a popular source of information to gather user opinions and reflections on a product or service. In this research the author analyses the user reviews of hotels using NLP and proves and suggest improved ways of analysing information using technologies such as Bidirectional Encoder Representations from Transformers(BERT). Performance of other NLP models are also measured and compared with different datasets. The data used to train and evaluate these methods are gathered from TripAdvisor which is considered as a good data source by researchers in this domain. Further the author also introduces a model of hotel aspect extraction from the reviews analysed.

Multiple experiments, examples and author's opinions are provided to prove and showcase the findings in this research. The results of the experiments show that BERT outperforms all other models and aspect extraction of hotels shows a relationship of the aspects considered by the customers and the hotel profile.