Project Work—Individual Reflection

Sorting sentiments of hotel reviews through machine learning

Group 8–01

Yap Hao Ming Darren (3i227, L)

2023

Contents

1 Individual Reflection

Embarking on a mathematics research project that delved into the world of sentiment analysis using machine learning techniques was both exhilarating and nerve-wracking for me, as someone who hadn't yet ventured into the realm of large-scale machine learning projects. I had only a rudimentary understanding of how machine learning was to be implemented into code, much less the mathematical theories behind them

that tied this machine learning endeavour with the project category—Category 8.

Our project revolved around using natural language processing (NLP) techniques to analyse the sentiments of hotel reviews, ranking them on a tripartite scale of positive, neutral and negative sentiment. After which, we would use the tokens in hotel reviews themselves to predict the overall sentiment of a hotel review, using machine learning models.

I was relatively unfamiliar with machine learning techniques upon starting the project, so I searched seemingly endlessly for similar examples conducted by tertiary researchers online to get a grasp of the fundamentals. Fortunately, I also sought the guidance of my peers who were more adept at machine learning, and consulted them on glitches and bugs within the machine learning algorithms I had implemented.

One of the major issues that we faced in the course of the project was during its conception. We knew we were going to focus on data analysis and statistical probability as our main area of research, but were not sure on what. After repeated consultations with our mentor, we decided to conduct the study on hotel reviews, since the post-COVID era has brought in a new wave of potential tourists from abroad. We had chanced upon a large dataset of ten thousand hotel reviews from international hotels in English from Kaggle, making the idea of analysing hotel reviews opportune.

By the end of the project, we had compared two models, the *random forest classifier* and the *logistic regression model*. By exploring the two different ways these models made predictions, we produced valuable insights regarding the models' precisions and accuracies.

In conclusion, despite the challenges and frustrations we faced in our project initially, we pulled through with insightful findings and a creative application of Mathematics in machine learning techniques. We had evaluated the reasons behind such findings carefully and learnt to exercise critical thinking and creative skills in this

project's development, making the project ever more fulfilling.