Subject

There are fake reviews for products and services online. Customers depend on genuine reviews to inform them for quality and economic and safety purposes. Businesses are paying to have more positive reviews about them online to increase sales and hotel stays to increase their profits.  Dubious performing businesses can damage the reputation of a platform for other transparent businesses. They tend to also generate revenue for that business which wouldn’t otherwise have been generated which raises ethical and legal concerns.

I’m interested in this topic because I’ve experienced it myself while travelling where a hotel has overly positive, gushing reviews and heavily doctored photos which give a false image of the business. I’ve found it extremely frustrating, and it can make me lose faith in a platform.

Hypothesis

AI can detect reviews which are not left by a genuine consumer using sentiment analysis and natural language processing

Research objectives

create an accurate machine model to detect if a review is fake or not

Data Sources

* As at least one of the data sources I’m planning on using outscraper.com. This website scrapes reviews from google, yelp, amazon and trip advisor and can output them into csv format
* I also intend to conduct a survey on the public to see the effect of fake reviews on a sample of consumers
* I intend to also contact hotels and local businesses to investigate have they been approached by such services and have they considered using them

Methodology

I intend to initially pre-process the data using basing lemmatization, stemming techniques. I intend to using advanced techniques such as Bag of words, semantics and sentiment Analysis, text classification and keyword extraction.

Preliminary literature review

['Fake' reviews on TripAdvisor a real problem for hoteliers – The Irish Times](https://www.irishtimes.com/business/technology/fake-reviews-on-tripadvisor-a-real-problem-for-hoteliers-1.607020)

<https://app.outscraper.com/tasks>