Enhancing the Accuracy of Inauthentic Review Detection using Machine Learning and Sentiment Analysis

Teresa Quain

A Thesis Submitted in Partial Fulfilment

of the requirements for the

Degree of

Master of Science in Data Analytics

Diagram

Description automatically generated with medium confidence

August 2022

September 2023

Supervisor: Vikas Tomer

Contents

[Foreword 1](#_Toc144929809)

[Problem Statement 2](#_Toc144929810)

[Introduction 3](#_Toc144929811)

[Methodology 31](#_Toc144929812)

[Dataset description 31](#_Toc144929813)

[Exploratory data analysis 34](#_Toc144929814)

[Data preprocessing 45](#_Toc144929815)

[Rule based classification 52](#_Toc144929816)

[Machine Learning 57](#_Toc144929817)

[Supervised Learning Method 57](#_Toc144929818)

[Unsupervised Learning Method 57](#_Toc144929819)

[Results and Evaluation 63](#_Toc144929820)

[Data pre-processing 63](#_Toc144929821)

[Rule-based classification 63](#_Toc144929822)

[Machine learning 66](#_Toc144929823)

[Supervised 66](#_Toc144929824)

[Unsupervised 68](#_Toc144929825)

[Conclusion 69](#_Toc144929826)

[References 71](#_Toc144929827)

[Appendices 80](#_Toc144929828)

[Appendix 1-Interview 80](#_Toc144929829)

[Appendix 3 Github Version Control 82](#_Toc144929830)

[Table 1 Summary statistics of dataset 36](#_Toc144929831)

[Table 2 Count of review likes 38](#_Toc144929832)

[Table 3 Top 10 frequent words 47](#_Toc144929833)

[Table 4 Top 10 weighted words 47](#_Toc144929834)

[Table 5 Sentiment summary statistics 50](#_Toc144929835)

[Table 6 Sentiment count 51](#_Toc144929836)

[Table 7 Results rule based classification 54](#_Toc144929837)

[Table 8 Cluster characteristics 58](#_Toc144929838)

[Table 9 Punctuation count per cluster 60](#_Toc144929839)

[Table 10 Top 10 reviewers ordered by review count 64](#_Toc144929840)

[Table 11 Top 5 restaurants order by punctuation count 64](#_Toc144929841)

[Table 12 Supervised ML parameters 66](#_Toc144929842)

[Table 13 Model results, Supervised ML 67](#_Toc144929843)

[Figure 1 Thesis structure flowchart 3](#_Toc144929844)

[Figure 2 Methodology structure 31](#_Toc144929845)

[Figure 3 Exploratory data analysis structure 34](#_Toc144929846)

[Figure 4 Histogram of Ratings distribution 37](#_Toc144929847)

[Figure 5 Histogram of review likes distribution 38](#_Toc144929848)

[Figure 6 Histogram of character count distribution 39](#_Toc144929849)

[Figure 7 Word count of frequent words 39](#_Toc144929850)

[Figure 8 Heatmap of variable correlation with Sentiment 41](#_Toc144929851)

[Figure 9 Heatmap of variable correlation with Review likes 42](#_Toc144929852)

[Figure 10 Count plot of restaurants vs. review rating 43](#_Toc144929853)

[Figure 11 Subplots of helpfulness, stop words, review likes & sentiment 44](#_Toc144929854)

[Figure 12 Data Preprocessing structure 45](#_Toc144929855)

[Figure 13 Review with displacy visualizer 48](#_Toc144929856)

[Figure 14 Named entity structure visualization 49](#_Toc144929857)

[Figure 15 Sentiment score vs. Rating 50](#_Toc144929858)

[Figure 16 Word cloud of positive reviews 51](#_Toc144929859)

[Figure 17 Word cloud of negative reviews 52](#_Toc144929860)

[Figure 18 Rule based classification structure 53](#_Toc144929861)

[Figure 19 Elbow method for optimal k 58](#_Toc144929862)

[Figure 20 Review likes vs Frequency per cluster 59](#_Toc144929863)

[Figure 21 Punctuation count vs Frequency per cluster 59](#_Toc144929864)

[Figure 22 Sentiment vs Frequency per cluster 60](#_Toc144929865)

[Figure 23 Level of detail vs Frequency per cluster 61](#_Toc144929866)

[Figure 24 Character count vs Frequency per cluster 62](#_Toc144929867)

[Figure 25 Character count distribution 65](#_Toc144929868)

# Foreword

The idea for my thesis came after some exploring around New Delhi, India, in which my partner and myself where scammed by false reviews of a supposed 4-star hotel with amazing reviews. After some more intense investigation it was quite obvious the reviews were fake, although the booking agent didn’t seem to have anyway of detecting this deception.

I would like to thank my loving Partner for the unwavering support and patience without which this undertaking would have been impossible.

I am extremely grateful to the College of Computing Technology and my supervisor Vikas Tomer, who’s knowledge and direction has been a great assistance on my journey.

# Problem Statement

As a result of this inauthentic’ reviews of products and services has become an epidemic online. Customers depend on genuine reviews to inform them for quality and economic and safety purposes.  These nuggets of wisdom and caution aim to bridge information and issues between buyers and sellers, or buyers and other buyers by providing information that may not be otherwise disclosed. Businesses are paying to have more positive reviews about them online to increase sales and hotel stays to increase their profits. Other reviews have been created by consumers who have been financially incentivised.  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.

Inauthentic reviews can lead to financial and safety risks for consumers as they may end up buying low-quality products or services or worse, be at risk of scams or fraud.

In addition to financial and safety concerns, inauthentic reviews also pose ethical and legal issues. Businesses that engage in fake review practices, such as paying for positive reviews or incentivizing consumers to leave reviews, are engaging in dishonest and manipulative behaviour that undermines the credibility of the review system. Furthermore, the presence of inauthentic reviews on a platform can harm the reputation of the platform and legitimate businesses that operate on it. Consumers may lose trust in the platform, leading to reduced traffic and revenue for both the platform and honest businesses. This can lead to a vicious cycle in which legitimate businesses are forced to compete with fake reviews to maintain their visibility, further eroding trust in the review system and the platform. Such practices also unfairly advantage some businesses over others, leading to an uneven playing field in the marketplace.

# Introduction

Natural language processing (NLP) or computational linguistics is one of the most important technologies of the information age. Applications of NLP are everywhere since now humans communicate almost everything via an online language: web searches, emails, language translation customer service, virtual agents, medical reports advertising and much more.

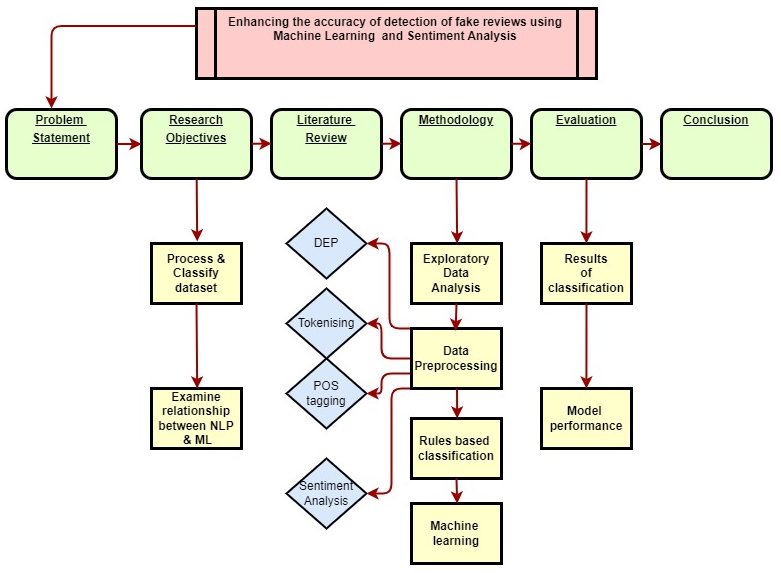


Figure Thesis structure flowchart

The structure of this thesis will follow fig 1. above:

The problem statement and ramifications of inauthentic reviews online will be outlined

Three research objectives will be presented to summarize the purposes of the study and organise the thesis into clearly defined components

The literature review and summary table will summarize the wide range of sources consulted for this paper and identify the gaps in existing knowledge

The methodology will be based on the code programmed in the Jupyter Notebook. It will consist of Exploratory Data Analysis (EDA) section, preprocessing using NLP techniques such as Dependency Parsing (DEP), Tokenisation, Part-of-Speech Tagging (POS) and sentiment analysis. Rule-based-Classification will be outlined and justified.

The Machine Learning (ML) will follow as an alternative to the Rule-based-Classification. This will include both supervised and unsupervised learning methods.

The Evaluation section will present the core findings of the Rule-based-Classification and ML modelling and describe how the outcomes were obtained and analysed

The conclusion will restate the original research objectives and present the evidence from the evaluation section supporting the findings. The entire paper will be summarized, and the key ideas discussed

In this thesis, the primary programming language that will be utilized is Python. Python is widely used in the field of data science and machine learning. It has an extensive collection of libraries and frameworks that can be utilized for NLP, natural language processing. Jupyter notebooks will be employed to write the code for the data cleaning as part of the pre-processing, the feature engineering, the sentiment analysis and the model training and critical evaluation. For NLP techniques, libraries such as NLTK (Natural Language Toolkit), scikit and TextBlob will be used for tasks such as tokenization, part-of-speech tagging, and named entity recognition. These libraries will also be used for feature engineering tasks such as extracting bag-of-words and tf-idf features.

Various machine learning algorithms will be utilized in this thesis proposal, including supervised and unsupervised learning algorithms, such as logistic regression, decision trees, and support vector machines (SVM) will be employed. The use of Python, Jupyter Notebooks, NLP techniques, and machine learning algorithms will be crucial in achieving the research objectives of this master's thesis. The combination of these tools and techniques will enable the development of an accurate and efficient algorithm for inauthentic review detection. Machine learning and sentiment analysis techniques can be leveraged to identify patterns of suspicious activity in review data, such as an unusually high number of positive reviews from a single IP address or similar language used across multiple reviews.

The colours of the various graphs, flowcharts and other plots are based on the google colour palette to reflect the dataset which consists of google reviews.

This topic is particularly interesting because many consumers have experienced it themselves while travelling, where a restaurant has a huge number of recently published, highly positive reviews which give a false positive image of the business. Consumers can find it extremely frustrating, and it may push a platform over the edge completely due to a lack of trust and a poor reputation. This manipulation of reviews can be dangerous, as it can mislead customers into making poor decisions that may have serious consequences. There is a good reason why customers are encouraged to be aware of this practice, and it is also important that they take steps to ensure that the reviews that they read are genuine.

**Keywords: Sentiment analysis; machine learning; e-commerce; natural language processing**

Hypothesis and Research Objectives

The hypothesis of this research is that AI can detect online reviews which are not left by a genuine consumer, by using natural language processing techniques and or machine learning. These falsified reviews are intended to generate additional business for a product or service. The enhanced and additional information, which this model would provide, would allow for more accurate financial projections of a business and promote a stronger customer base. Moreover, it would also enhance a company's reputation online and create more opportunities to conduct targeted marketing campaigns and increase sales.

Through this analysis, the following research objectives will be pursued:

To process and classify a text dataset in depth of online reviews from restaurants in Ireland using NLP techniques such as POS tagging, entity identification and semantic analysis

To implement a rule-based classification system to detect if online reviews left are completed by authentic patrons.

To generate several machine learning models on the trained dataset and compare their performance results and overall effectiveness with the rule-based classification system.

Literature Review

A literature review has been conducted, followed by the presentation of the proposed problem of detecting inauthentic reviews posted online which have not been left by a genuine consumer.

In this literature review the focus was aimed at reviewing a wide range of literature related to the business of online reviews for products and services. It was necessary to consult a diverse set of sources for this project, including various online journalistic articles and research studies, and many of these are listed below, all of which been cited and can be found in the references section. By consulting a diverse range of sources, including research studies and online journalistic articles, this literature review will provide a comprehensive overview of the subject and lay the groundwork for the proposed research project.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source, Title,**  **Author** | **Source Type** | **Methodology** | **Performance**  **Metrics** | **Critique Remarks against RO's** |
| [1] How to spot fake  reviews on Amazon, Best Buy, Walmart and other sites, Rick Broida | journal article cnet | internet research | wide, in-depth assessment on other resources | useful assessment of  existing methods |
| [2] FakeSpot, private  company, Khalifah | browser extension | NLP + AI | unclear, no stats | no detail on DA  used |
| [3] FTC Puts Hundreds of  Businesses on Notice  about Fake Reviews and  Other Misleading Endorsements, Federal Trade Commission | press release | penalty notice offence issued | undisclosed | shows prevalence of  inauthentic reviews globally |
| [4] amazon-fake-reviews-facebook-groups, Vish Gain | article silicon republic | Identify + target facebook groups who are responsible | no stats, just comments on other methods | useful knowledge of existing social media groups |
| [5] Inside the Underground  Market for Fake Amazon  Reviews, RAJVARDHAN OAK | article, wired magazine | survey on prevalence of people completing fake reviews for refunds | no stats, just comments on other methods | useful knowledge of existing groups |
| [6,7] All You Need to Know  on Fake Reviews and  False Ratings, ECC Ireland | study report | detail business testimonials and review aggregators | free product reimbursement, pillow 20$ | valuable legal background information on this issue. Complete list of companies included |
| [8] Why we usually can't tell when a review is fake, Shabnam Azimi and Alexander Krasnikov of Loyola University of Chicago and Kwong Chan of Northeastern University | study report | participants separated real review from fake ones depending on their on their opinion | no metrics published. Study suggests were more like to believe negative review | no results published, methodology interesting, valuable methods to incorporated into rule classification |
| [9] Fake Review Watch, Kay Dean | website | investigating user  accounts | unclear, no stats | no DA used, manual research |
| [10] Method to Facilitate E-Commerce Buying Power by Using Machine Learning Techniques, Junzhi Liu | research article, ResearchGate | Naïve Bayes and Logistic Regression, word clouds | delivery time for commerce | valuable usage of word clouds +other visualisations |
| [11] Illusions of truth—Experimental insights  into human and algorithmic detections  of fake online reviews, Daria Plotkina , Andreas Munzel , Jessie Pallud | research paper,  Science Direct | linguistic analysis tool - Coh-Metrix | unclear, no stats | wide range of techniques used. Valuable paper for research |
| [12] Review Meta, private company | website, app , browser  extension | Independent testing to give user an 'Adjusted Rating' | website claims to analyse 1M + /day | not enough detail on algorithm employed. Valuable resource for dataset on fake reviews |
| [13] A Fake Review Detection System Using  NLP and Machine Learning Techniques, P. Aishwarya Sri, R. Vamshidhar Reddy | research paper, international journal of Scientific & Engineering Research | Supervised + Unsupervised learning, NLP, decision tree classifier, a rule-based classifier, Naïve Bayes | unclear, no stats | apart from lack of results a relevant paper for my work |
| [14] Amazon has removed hundreds of thousands of incentivized reviews since it banned the practice, Rob Thubron | techspot, journal article | Review Meta, 5 million reviews across 32,060 products | unclear, no stats | no detail on DA  used, relevant background information |
| [15] Inside the War on Fake Consumer Reviews, MEGAN MCCLUSKEY | Time magazine, article | how to inform public, consumer advocate | no metrics, theoretical study | useful knowledge of current level of issue |
| [16] Amazon continues to take action against fake review brokers, Amazon staff | Amazon policy news | industry-leading tools to detect and block fake reviews | 2022- Amazon took legal action against 90 bad actors who facilitated and solicited fake reviews | relevant background information, active team in place from Amazon side |
|  |  |  |  |  |

**Fakespot** [2], this can be added as a browser extension or a mobile app and claims to provide secure shopping on amazon, Sephora, eBay and Walmart. Fakespot claims to ‘protect you from getting ripped off when shopping online’ and will ‘get the truth about products, reviews and sellers before you buy’. The website does not provide any technical details on how exactly it is filtering out certain reviews or how it calculates its own star rating for products. It simply claims that it uses ‘AI to detect fraudulent product reviews and third-party sellers in real-time’. Suspiciously the website also contains plenty of five-star positive reviews. They also offer a package for businesses called ‘Trust AI’ which claims to be a powerhouse NLP AI which will work across textual content and extract valuable intelligence from consumers. They have written that their product will find inauthentic accounts and bots on your service for fast and easy removal. The product also contains a review sentiment which will apparently allow you to uncover trends and insights from your customer reviews. Overall, the site promises a lot but does not provide any real evidence to back it up. The browser extension does function and tends to give a slightly lower rating that the shopping website, so it is filtering out certain reviews (either real or inauthentic).

**[4]** There are many new and business platforms which have thrown their two-cents worth in the ring**]. Silicon Republic** is an American sci-fi news platform. Vish Gain, a journalist there was the author of a published article in July 2022 which describes a court case led by Amazon suing 10,000 Facebook groups over inauthentic reviews. Amazon was claiming that these groups orchestrate inauthentic product reviews on the website in exchange for money or free products. Amazon claimed at the time, the information would help to ‘complement the technology and continuous monitoring’ that it already did to reduce inauthentic reviews. They requested other social media companies such as Meta, inauthentic book’s parent company to step in at the time to support the case. One of the groups in particular names as ‘Amazon product Review’ had 43,000 members. This article unfortunately doesn’t give much more detailed information about the numbers of reviews involved and the site has not provided an update on this since which makes it difficult to decide which side of the fence, the platform is sitting on.

**ReviewMeta** [12] is another similar site that works only with Amazon, but worldwide. They position themselves as an **independent** checker tool that analyses reviews and ‘helps your shopping experience’. It also comes in the form of a mobile app, a browser extension. Under the ‘How it works’ section are a series of short entertaining video’s, a link to a podcast with the creator (Tommy Noonan), a nice little background story of how he apparently came up with the idea for the website. The site is positively geared towards Amazon and **blames brands** for choosing to ‘abuse the platforms that were created to help customers, flooding them with low quality and biased reviews’ which it says the brands do to try and boost their own profits. Like ‘fakespot’ they claim to be performing some statistical modelling to present a truer picture of an online product. The ‘report card’ which the website can generate from an amazon link judges reviewers for their average rating, their word count, repetitive phrases and review history. The analysis this website seems to perform is interesting, and the transparency is commendable, it does seem biased towards Amazon though they do claim to be completely independent.

There are of course legal consequences for those who choose to commit what may seem like a petty crime.

[13] **‘A Fake Review Detection System Using NLP and Machine Learning Techniques’** by P. Aishwarya Sri and R. Vamshidhar Reddy describes a method of creating a spam review detection model and ensuring a high level of accuracy through ML methods. Detection feature methods are dived into two categories ‘review centric’ and ‘reviewer centric’, that is detection methods based on a single review or detection methods based on all reviews from a particular reviewer respectively. Pre-processing techniques include standard NLP methods such as stemming, LC conversion, Tokenisation and stop word removal to ensure the dataset sourced from Amazon reviews was as clean and functional as possible. A combination of unigrams, bi-grams and tri-grams were applied to the review information to ensure the adjectives were being grouped correctly with the appropriate nouns as implied by the reviewer. Word embedding allowed the words to be classified into low dimensional vectors.

An in-depth approach was taken by the author through machine learning. Supervised methods such as Decision tree classifier, Bayesian system, Naive Bayes and rule-based classifiers were executed. Rule-based Classifiers [13] were applied as a method of flagging reviews as ‘authentic’ or ‘inauthentic’. Since there is no ‘one size fits all’ rule, a combination of approaches was applied to catch as many indicators as possible. Unsupervised learning methods such as twice-clustering and K-means clustering were also applied. They were justified by their learning strategy of ‘examining the relationship between variables’ which can be suitable for unlabelled datasets. Overall, the article is quite simplistic, but it is well written and straightforward. The techniques outlined are relevant to this paper’s own research, in particular the supervised machine learning methods.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source, Title, Author** | **Source Type** | **Methodology** | **Performance Metrics** | **Critique Remarks against RO's** |
| [17] Differential Query Semantic Analysis: Discovery of Explicit Interpretable Knowledge from E-Com Search Logs, Sahiti Labhishetty,  Cheng Xiang Zha,  Min Xie, Lin Gong | E-Com search logs | limited to data format | bridging the vocabulary gap, comparative analysis of search intent, and alleviation of the problem of tail queries and products. | novel strategy using nlp techniques. Limited access to full article |
| [18] Fake review identification and utility evaluation model using machine learning , Wonil Choi, Kyungmin Nam, Minwoo Park1, Seoyi Yang | Corpus-collection of linguistic data | requirement to dig further to read other articles | N/A- not published | Good summary overview of techniques availability. Doesn’t contain in-depth knowledge |
| [19] The Indian repository of resources for language technology, Narayan Choudhary | fifty people and involving thousands of resource persons, covering twenty major languages of India | human resource needed for each language | datasets in 20 scheduled languages of India so far collected | these works are crucial in the development of language technology for Indian languages |
| [20] Natural Language Processing (Almost) from Scratch, Ronan Collobert, Jason Weston, Leon | large unlabelled data sets (∼ 852 million words) | no task-specific engineering | State-of-the-art systems on four NLP tasks. Performances of between 70-90% | wide range of techniques used. Valuable paper for research |
| [21] Commonsense reasoning and commonsense knowledge in artificial intelligence, Ernest Davis, Gary Marcus, NYE | commentary no experiment | e.g. Google translate, hand coding required | N/A not published | insightful article, innovative techniques used to imitate common sense |
| [22] Searching Better Architectures for Neural Machine Translation, Yang Fan; Fei Tian; Yingce Xia; Tao Qin; Xiang-Yang Li; Tie-Yan Liu | translation via audio, speech and language processing | resource requirement | permeance model scores | difference approach, not useful to this thesis but still interesting |
| [23] Sentiment analysis using product review data, Xing Fang & Justin Zhan | Stanford Sentiment 140 Tweet Corpus | quality of the opinions on Twitter, truth of such online data is not always available. | Performance based on F1 score | valuable legal background information on this issue. Complete list of companies included |
| [24] GA, MR, FFNN, PNN and GMM based models for automatic text summarization, Mohamed Abdel Fattah, Fuji Ren | 200 Arabic articles in the domain of politics and 150 English articles in the domain of religion from internet archives | semantics such as synonymy, polysemy, and term dependency not taken into account | average nr. Of sentences compressed/min per model type | only baseline approach results included |
| [25] DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization, Authors:  Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, Karrie G. Karahalios | commentary no experiment | dropdown menu. Needs manual correction from user | dependent on subject | interesting adaption for visualisations |
| [26] Advances in Automatic Text Summarization, Inderjeet Mani and Mark T. Maybury | collection of the most important writings in automatic text summarization | produce an abridged version for a particular user or task | N/A not published | different approaches discussed |
| [27] Natural language processing: state of the art, current trends and challenges, Diksha Khurana, Aditya Koli, Kiran Khatter | Corpus-collection of linguistic data | no research experiment included. General overview | no research experiment included. General overview | no research experiment included. General overview |
| [28] Fake online reviews cost $152 billion a year. Here's how e-commerce sites can stop them, Jonathan Marciano | official figures and self-reporting by e-commerce sites (including Trip Advisor, Yelp, TrustPilot and Amazon | Cost of court cases | direct influence of fake online reviews on global online spending is $152 billion | no technical description of tools available, only overview included |
| [29] Text Generation, Kathleen McKeown | Corpus-collection of linguistic data | computational hardware | human judgement | no quantitative performance metric |
| [30] How to implement CNN for NLP tasks like Sentence Classification,  Rajat Newatia | Sentiment Labelled Sentences Data Set from the UCI Machine Learning Repository | nr. Of training samples, hyper parameter tuning | accuracy of CNN model performance | excellent paper with detailed description on code employed |

The article ‘Fake review identification and utility evaluation model using machine learning’ provides some in- depth reading on the subject mentioned in the title [18]. The article was published in January 2023 on the frontier’s platform. The article begins by reiterating the need to monitor fake reviews and highlights some of the consequences of leaving them unchecked. Some of these include platforms which are on the verge of losing credibility and traffic and a decline in sales by other affected vendors. The authors warn that with the general rise in online shopping transactions, the reliance by customers on reviews in making purchasing decisions has also increased dramatically in recent years. The data was sourced through ‘selenium on the **Korean Naver shopping mall platform**. The authors explain that this platform would provide a good representation of review posts and the platform collects reviews from other shopping mall platforms and just republishes them. The aim of the study is to propose an algorithm that utilizes machine learning to firstly remove macroscopic views of fake reviews and to place reviews that provide useful information to buyers at the top of displayed platform results to provide a better experience for all platform participants. A wide variety of data analysis techniques **Neural networks, Generative Pre-Trained transformers and Bidirectional Encoders** were employed in this study. These techniques use the benefits of deep learning and machine translation. Each technique is explained such as the Support Vector Machine (SVM), which was used in this case to define decision boundaries and classify unclassified points, along those boundaries. Several natural language techniques and libraries were also used which may serve as a valuable support for this thesis. Libraries such as BERT and GPT are valuable transformer encoders. Some downfalls of the sentiment analysis are mentioned which shows the authors transparency, such as the adverbs not being included which influence the intensity of a word.

One of the interesting natural language processing techniques which the authors described was tokenizing the emotion analysis data using KoNLPy’s OKT. The reason for using Okt was its function of automatically correcting typos. This can have a great effect on improving accuracy when applied to review data that would have many typos. Regarding clustering, Hierarchical Clustering technique showed the most meaningful results in this study. SVC, LGBM, logical regress and KNN models performed during the supervised learning showed high accuracy with high AUC score values in determining inauthentic models. These results were verified by performing in-depth interviews with customers in this case for a tooth whitening product.

The significance of this study is in its link to from data analysis to actual mobile commerce **sales performance**. The study confirms quantitative and non-quantitative characteristics of customer reviews referred to by customers for decision-making has effect on the financial profits of an enterprise. This bold statement could definitely encourage more interest and investment on the part of businesses to take more of a serious attitude to their online profile. The wish of the authors by undertaking this study is that e-commerce platforms increase service reliability, prevent information overload, prevent participants from leaving the platform, and provide better online shopping experiences to attract more consumers by detecting and filtering out fake reviews using machine learning algorithms.

The paper titled ‘Natural language processing: state of the art, current trends and challenges’ , written by [Diksha Khurana](https://link.springer.com/article/10.1007/s11042-022-13428-4#auth-Diksha-Khurana), [Aditya, Koli](https://link.springer.com/article/10.1007/s11042-022-13428-4#auth-Aditya-Koli), [Kiran Khatter](https://link.springer.com/article/10.1007/s11042-022-13428-4#auth-Kiran-Khatter) & [Sukhdev Singh](https://link.springer.com/article/10.1007/s11042-022-13428-4#auth-Sukhdev-Singh) and published in 2022 presents a fascinating insight into the advances in NLP in recent years [20]. It has spread into areas of information extraction, medial fields, cyber security and summarization, and is of interest to many professionals such as linguistics, psychologists and philosophers. The authors define it as a ‘tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages. The paper describes many libraries and applications of NLP which is very interesting for the research of this thesis.

Neural networks are mentioned for example as having revolutionized the field of Natural Language Processing (NLP) by enabling the handling of variable length inputs, which is particularly useful for processing text. The introduction of neural networks brought a significant change in the way NLP problems were approached and solved. For instance, sequence-to-sequence mapping framework is a general approach for mapping sequences of variable length to another sequence of variable length, and has been applied to various NLP tasks such as machine translation, summarization, and question-answering. In this approach, the encoder network processes the input sequence and generates a fixed-length vector representation, which is then used by the decoder network to generate the output sequence. The use of neural networks has led to significant improvements in NLP performance on various tasks, such as sentiment analysis, named entity recognition, and text classification. Moreover, the development of novel neural network architectures such as Convolutional Neural Networks (CNNs) and Transformers has further improved the state of the art in NLP. CNN’s have also been employed in sentiment analysis, classification and machine translation.

The authors give a valuable summary on the process of sentiment analysis. It tends to involve the use of two linguistic resources: the sentiment lexicon and the sentiment pattern database. The sentiment lexicon is used to identify positive and negative words in the documents, and the sentiment pattern database is used to analyse the sentiment patterns in the text. The analysis assigns ratings on a scale of -5 to +5 based on the identified positive and negative words. The currently used tag sets for sentiment analysis are mainly derived from English and designed for Indo-European languages. However, research on Asian and Middle Eastern languages is relatively limited in comparison and may benefit from further research in the future. Sentiment analysis however can also be quite limited, the author’s comment how online, when for example slang is used, it is necessary to consider noun phrases. They use the example of tweets and the Named entity recognition (NER) technique which is used to recognize and separate and aid in classifying text into predefined classes. Considering these metrics in mind, it is necessary to evaluate the performance of an NLP model for a particular application and a particular case.

The reference list for this paper is extremely in-depth and includes books, papers, google scholar articles and executed studies involving the application of NLP techniques on real world problems. The paper focuses not only on various approaches and evaluation metrics of NLP, but also the history of the area and recent developments in literature.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source, Title, Author** | **Source Type** | **Methodology** | **Performance Metrics** | **Critique Remarks against RO's** |
| [31] Using Natural Language Processing and Network Analysis to Develop a Conceptual Framework for Medication Therapy Management Research, | publication, Towards Data Science | NLTK package python | feature extraction model performance | methods not fully justified |
| [32] Machine Translation: Past, Present, Future, William John Hutchins | research article,  peerj.com | language parsers and emoticons, context-level sentiment analysis techniques, pre-processing methods, and lexical resources | machine + deep learning model performance | detailed article, in depth literature review |
| [33] Sentiment Analysis on Amazon Reviews, Enes Gokce | research article, Towards Data Science | Recurrent Neural Network | no specified | no performance metric detailed |
| [34] Sentiment analysis techniques, challenges, and opportunities: Urdu language-based analytical study, Muhammad Irzam Liaqat1, Muhammad Awais Hassan​1, Muhammad Shoaib1, Syed Khaldoon Khurshid1, Mohamed A. Shamseldin2 | research article, Digital Library | new language model to simultaneously cluster and summarize documents by making use of both the document-term and sentence-term matrices | Experiment on interpretability of the generated summaries. | innovative modelling method used, valuable for data collections when grouping is required |
| [35] Recurrent Neural Networks and Natural Language Processing, Christopher Thomas BSc Hons. MIAP | publication, Science Direct | applying meaning representation language (MRL) that facilitates the uniform interpretation | ACM model performance | Paper doesn’t go into enough detail to describe construct of rules and techniques |
| [36] Integrating Document Clustering and Multi document Summarization, Dingding Wang,  Shenghuo Zhu,  Tao Li,  Yun Chi,  Yihong Gong | publication, Wired journal | facebook groups offering free refunds for reviews posted | salary level of successful review agents | In-depth research by journalist into inauthentic review groups organisation. Reference to similar online services |
| [37] Semantics and Quantification in Natural Language Question Answering, W.A. Woods | research article, Journal of Scientific + Engineering Research | NLP, feature extraction, training, classification | K means clustering algorithm until convergence is achieved; ideal classifier identified | high level of detail on machine learning algorithms, performance evaluation not included |
| [38] Inside the Underground Market for Fake Amazon Reviews, RAJVARDHAN OAK | research article | problems are formulated as an integer linear program (ILP) and solved using public domain software. | Model performance of both methods presented, ACM | no supervised methods explored |
| [39] A Fake Review Detection System Using  NLP and Machine Learning Techniques, P. Aishwarya Sri, R. Vamshidhar Reddy | research article, Springer Link | memory-based anti-spam filtering | memory-based filter model performance | detailed article, in depth literature review |
| [40] Global unsupervised models for keyphrase based meeting summarization, Korbinian Riedhammer, | book | combination of approaches inc. from neuropsychology, psycholinguistics, and artificial intelligence | context of utterance to determine the proper meanings of words and sentences | broad based approach from various angles, deep understanding of the topic |
| [41] A Memory-Based Approach to Anti-Spam Filtering for Mailing Lists, Georgios Sakkis, Ion Androutsopoulos, Georgios Paliouras, Vangelis Karkaletsis, Constantine D. Spyropoulos & Panagiotis Stamatopoulos | Article, Research Gate | machine learning algorithms to build the classifiers | feature designing and fake review detection models | valuable article for own research. Innovative distinctions between types of fake reviews |
| [42] LEXICAL AMBIGUITY RESOLUTION,  Steven Small, Publisher: Morgan Kaufmann | Article, Research Gate | meta-graph to construct a heterogeneous information network | spamming score per group, overall detection performance, Comparison of P@ k for five methods | logical methodology flow detailed well in graphics |
| [43] Survey on Fake Review Detection Research, L.-Y. Li, B. Qin, T. Liu | Article, Research Gate | Word2Vec model, N nearest neighbour user relationship, DBSCAN algorithm | Experimental results of detection performance | approach not always justified or explained. Good model performance |
| [44] Network Embedding-Based Approach for Detecting Collusive Spamming Groups on E-Commerce Platforms, Jinbo Chao, Chunhui Zhao, Fuzhi Zhang | Article, Research Gate, Department of Telematic Engineering Systems | review scraper, feature computation, fake classifier | 82% F-Score on the classification task, Ada Boost classiﬁer has been proven to be the statistical means according to the Friedman test. | good summary tables, clear framework table |
| [45] Detecting review spammer groups based on generative adversarial networks, Jinbo Chao, Fuzhi Zhang | research study | HanLP package for translation, Latent Dirichlet Allocation (LDA) method for topic modeling, pyLDAvis for modelling | perplexity value and coherence score | interesting approach, no mention of fact that reviews could be false and basing data off them could be biased |
| [46] A framework for fake review detection in online consumer electronics retailers, Rodrigo Barbado, Oscar Araque, Carlos Iglesias | research study, Springer article | Convolution neural network, multidimensional feature representation is used to classify reviews | CNN performances of 0.988, 0.987, and 0.994 | valuable article for own research. Innovative hybrid approach for text, emotions and ratings |
| [47] How Could Consumers’ Online Review Help Improve Product Design Strategy? Wei Miao, Kai-Chieh Lin , Chih-Fu Wu , Jie Sun , Weibo Sun , Wei Wei and Chao Gu | research article, research ate | Algorithm-Based Filtering | recognition accuracy rate based of punctuation score, sentiment score, vocabulary score | excellent visual flowcharts for rule-based classification |
| [48] DHMFRD – TER: a deep hybrid model for fake review detection incorporating review texts, emotions, and ratings, Ramadhani Ally Duma, Zhendong Niu, Ally Nyamawe, Jude Tchaye-Kondi | research article, research gate | natural language processing; Polish language; machine learning; random forest | F1 score of 0.92 and 0.74 when detecting fake accounts and reviews | large dataset, novel metrics to detect fake review, good summary in lit review |

The article, ‘Research on False Review Detection Methods: A state-of-the-art review [49] written by [Arvind Mewada](https://www.researchgate.net/profile/Arvind-Mewada) and [RUPESH KUMAR DEWANG](https://www.researchgate.net/profile/Rupesh-Dewang) and published on the research gate platform provides an extremely in depth review into modern analytical methods for detecting fake reviews. The article also contains extremely well laid out methodology through the use of flow charts and tables. The authors have divided the content of spam reviews into 3 categories: untruthful reviews, reviews on brands only and non-reviews. By ‘non-reviews’ the authors mean the review contains no useful information specific information and only generic comments. Reviews were sourced from Yelp website, Tripadvisor Reviews and Amazon book and product reviews. The authors employ feature engineering to detect the clues left by reviewers in fake reviews through lexical, sentiment and syntactic methods. A mathematical model was established to quantify the characteristic values of these behaviours.

The authors have used rule-based classification to score the reviews based on certain characteristics such as their rating, rating consistency, time frame between reviews and product launches and the percentage of sentiment values. Other methods included linguistic NLP methods such as Bag of Words, POS tagging, Word count features, Grammatical syntactic structure analysis. The discussion and results section are well structured, classifiers and machine learning methods are listed such as SVM, Random Forest. A few possible future improvements are listed such as optimising the labelled data, however the authors do not include why these were not included in this article. Each dataset, method, outcome matrix and model results is listed clearly in Table 5. The article limits itself by only using false review text as the identification target, however future directions are laid out in the conclusion section towards future research in the field of false reviews and false reviewer groups.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source, Title, Author** | **Source Type** | **Methodology** | **Performance Metrics** | **Critique Remarks against RO's** |
| [49] Research on False Review Detection Methods: A state-of-the-art review, Arvind Mewada, [RUPESH KUMAR DEWANG](https://www.researchgate.net/profile/Rupesh-Dewang?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIiwicGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIn19) | research article, research gate | Algorithm-Based Filtering | recognition accuracy rate based of punctuation score, sentiment score, vocabulary score | excellent visual flowcharts for rule-based classification |
| [50] Detecting Fake Reviews in Google Maps—A Case Study, Paweł Gryka, Artur Janicki | research article, research gate | natural language processing; Polish language; machine learning; random forest | F1 score of 0.92 and 0.74 when detecting fake accounts and reviews | large dataset, novel metrics to detect fake review, good summary in lit review |
| [51] Aspect‑based classiﬁcation method for review spam detection, Mengsi Cai, Yonghao | research study, Springer article | Bi-LSTM model to automatically extract massive aspect words which are clustered into diﬀerent aspect categories by the K-means algorithm | review spam detection by about 16.11% to 38.86% compared with textual and behaviour features | dataset size not detailed; some assumptions made about spam reviewers |
| [52] Expected and materialised information source use by municipal officials: intertwining with task complexity, Miamaria Saastamoinen | research article, research gate | analyse the use of information sources in the context of varying task complexity and from the perspective of task performers | Mann-Whitney statistically significant, p=0.03 between simple/complex tasks | slightly confusing language, well presented table of results |
| [53] A comprehensive survey of various methods in opinion spam detection, Arvind Mewada | RUPESH KUMAR DEWANG | . Machine learning methods and natural language processing techniques | quantitative comparison to aspects of feature design, model methods, datasets, and rating indicators | clear distinction f types of review eg. Fake/brand/non review/ spammer. No innovative ML modelling used |
| [54] Sentiment Analysis for E-commerce Product Reviews: Current Trends and Future Directions, Salma Elzeheiry, Wael Ali Mohammed Gab Allah, Nagham Mekky, Mohammed Elmogy | research article, research- gate | word2vec, deep learning, sentiment analysis, bigram, and Glove, TF-IDF | Area under AUC , ROC curve | clear methodology flowcharts |
| [55] Method to Facilitate E-Commerce Buying Power by Using Machine Learning Techniques, Junzhi Liu | research article, research gate | Naïve Bayes and Logistic Regression, word clouds | delivery time for commerce | valuable usage of word clouds |
| [56] Whose reviews are most valuable for predicting the default risk of peer‑to‑peer lending platforms?, Liting Li, Haichao Zheng, Dongyu Chen, Bin Zhu | research article, Electronic Commerce Research (2022) | linear regressions of the number of manipulated reviews / average rating on transaction volume. | correlation value between platform duration and manipulated positive review count | Limited to Chinese dataset |
| [57] Fake Review Detection: Classification and Analysis of Real  and Pseudo Reviews, Arjun Mukherjee , Vivek Venkataraman , Bing Liu , Natalie Glance | research paper, Google scholar | using the information theoretic measure KL-divergence and its asymmetric  property on 2 types of reviews (real/fake) | classification  accuracy of 89.6% with bigram features | interesting psycholinguistic  phenomena about forced and natural fake reviewers |
| [58] A deep learning approach for detecting fake reviewers, Dong Zhang, Wenwen Li, Baozhuang Niu, Chong Wu | research paper, Science direct | Behaviour sensitive feature extractor, context-aware attention mechanism, and fake reviewer detection | long delivery time with deep learning | novel deep learning-based framework for detecting suspicious reviewers |
| [59] Creating and detecting fake reviews of online products, Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-gyo Jung, Bernard J. Jansen | research paper, Science direct | 2 models CPT-2 and ULMFit represent different types of NLP architectures | performance – accuracy of 63% with text features and 78% with all available features | well-designed visualizations |
| [60] An Approach for Detecting Spam in  Arabic Opinion Reviews, Ahmad S. J. Abu Hammad | Msc. Thesis | classification method to find a way to classify the  Arabic opinion reviews, whether it should be spam and non-spam | F-measure 99.6% | well laid out thesis, quite basic coding used |
| [61] Distortion as a Validation Criterion in the Identification of Suspicious Reviews, Guangyu Wu, Derek Greene, Barry Smyth, Pádraig Cunningham | research paper, UCD | an assessment of shill detection mechanisms on a dataset of hotel reviews | CPS Score, level of distortion, PPS score | valuable paper for own research, well written, in-depth discussion |
| [62] Removing order effects from human-classified datasets: A machine learning method to improve decision making systems, Dmitry Romanov , Valentin Molokanov , Nikolai Kazantsev , Ashish Kumar Jha | research article, Science direct | basic classification algorithms Naïve Bayes to test our model's efficacy in eliminating the order effect | Naïve Bayes significance level | paper fails to answer how to remove other forms of cognitive bias, such as gender and racial biases. |
| [63] A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information, Dong Zhang , Wenwen Li , Baozhuang Niu , Chong Wu | research article, Science direct | using 133 unique features from the combination of content and behaviour-based features to detect fake reviews using ML classifiers. | results (accuracy, precision, recall and F1) from Scikit library, 7.73% for the fake class and 99.3% for the genuine class using a Multilayer Perceptron classifier | well discussed, innovative hybrid approach. Results clearly presented |
| [64] Consumers rule: How consumer reviews influence perceived trustworthiness of online stores, Sonja Utz, Peter Kerkhof, Joost van den Bos | research article, Science direct | effects of online reviews on perceived trustworthiness of an online store | opinions of participants document in table | similar experiments out there, nothing innovative about this experiment |
| [65] The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews, Do-Hyung Park , Sara Kim | research article, Science direct | experiment asking participants with varying level of expertise about multimedia product | no significant differences in gender (F(9, 240) = 0.766, p < 0.648), age (F(9, 240) = 0.862, p < 0.560), and frequencies of online purchase (F(9, 240) = 0.896, p < 0.530) | study makes several theoretical contributions |
| [66] Suspicion of online product reviews as fake: Cues and consequences, L. Jean Harrison-Walker, Ying Jiang | research article, Science direct | study of types of cues used by customers to determine if review is authentic | findings discussed not evaluated quantitatively | good insight for brands and market research, useful cues to consider |
| [67] Understanding online fake review production strategies, Snehasish Banerjee , Alton Y.K. Chua | research article, Science direct | participants outline strategy on method, challenges faced, demographic details | planning style, knowledge style score and creating style score values to explore mapping between production strategy and cognitive style | Nice use of Anova to differentiate between clusters. Well thought out performance evaluation |
| [68] The more they know: Using transparent online communication to combat fake online reviews, Yiru Wang , César Zamudio , Robert D. Jewell | research article, Science direct | communicating the platforms’ actions online in transparent manner | four-stage framework, VIEW (Verify, Inform, Explore, and Watch), provides a roadmap for online review platform managers | theoretical discussion, missing concrete performance results |
| [69] Mind the fake reviews! Protecting consumers from deception through persuasion knowledge acquisition, Murilo Costa Filho , Diego Nogueira Rafael | research article, Science direct | pilot study with consumers against an original hypothesis | results suggest fake reviews are much more likely to go unnoticed by consumers when they are not equipped with tools to detect them | well written theoretical discussion, support for more transparency with consumers |
| [70] Unfolding the characteristics of incentivized online reviews, Ana Costa , João Guerreiro , Sérgio Moro , Roberto Henriques | research article, Science direct | data mining approach to predict whether or not a new review published was incentivized based on several review features such as the overall rating, the helpfulness rate, and the review length | word count, number of sentences, number of characters, Kruskal-Wallis test | interesting conclusion that incentivised reviews have longer character count and can be used as deciding factor |

**To conclude this literature review.** The continued media focus on inauthentic reviews, driven in part by such websites as Review Meta [4] and Fake spot’s [3], their relentless publicity drive is taking attention away from much more serious issues that the media does not cover, such as click farming, book stuffing, incentivized purchasing, and mass gifting however this goes beyond the subject of this project thesis. The article published on the frontier’s platform [17] is focused on a Korean shopping platform. The author has undertaken a very deep look at natural language techniques combined with several machine learning models which is a valuable piece of research for this thesis. The study demonstrates a strong link between sales data for a particular product and the presence of online fake reviews for that product. This information is then backed up by in-depth interviews from platform participants making overall a very strong case. [12] This report from Amazon shows they acknowledge how widespread the issue is, simply by the fact that they claim to have over 12,000 employees working on the issue and that they have filed six lawsuits against companies in 2023 for fraud and abuse of the platform. The article in Wired provides and interesting insight into the multiple social media groups that are being financially incentivised to purchased targeted items and the write a review about it. Awareness of the darker side to these reviews has widened the research of this thesis greatly and has provided valuable opinions which contribute to the depth of this report. The knowledge of the existence of such reviews has broadened the research of this thesis, providing valuable insights and opinions that contribute to the depth of this report. Therefore, it is crucial to develop accurate methods for detecting fake reviews and ensuring the integrity of review systems to protect consumers and businesses' interests.

# Methodology

The analysis will be set out in the following format as seen in fig 2. below. This methodology will describe the initial dataset, the EDA, exploratory data analysis that was performed, the preprocessing necessary for this dataset, the rules-based classification system based on predefined NLP methods and a comparison to alternative machine learning methods.

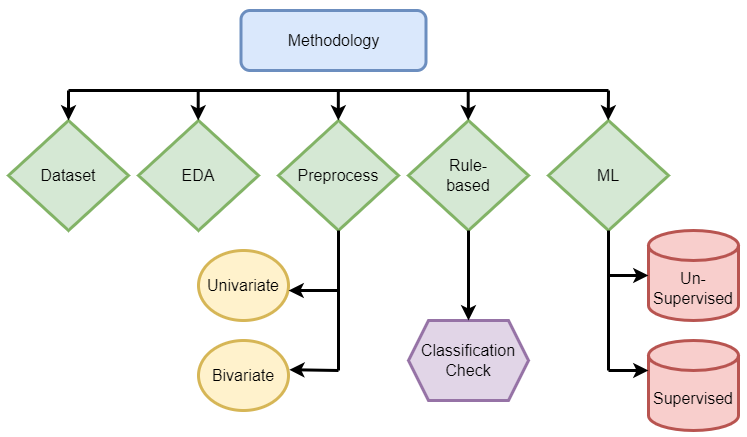


Figure Methodology structure

## Dataset description

The chosen dataset was sourced form outscraper.com. This dataset was chosen as it allows free access to scraped google reviews and contains valuable information about both the review and the reviewer. It consists of 26 data columns and 1328 data rows or google maps reviews using the queries with ‘Dublin’ and ‘Galway’ as location and ‘restaurant’ as a filter between February 2016 and April 2023. The csv data was read into a panda’s dataframe.

The dataset contains 8 variables with qualitative data such as:

* ‘reviews’,
* ‘rating’
* ‘author id’
* ‘owner\_answer\_timestamp’
* ‘review\_rating’
* ‘review\_timestamp’
* ’ review\_likes’
* ‘review\_id’,

and 17 variables with quantitative information such as:

‘query’

‘names’

‘google\_id’

* ‘place\_id’
* ‘location link’
* ‘review\_per\_score’,
* ‘review\_id’
* ‘author\_link’
* ‘author\_title’
* ‘author\_image’
* ‘review\_text’
* ‘review\_img\_url’
* ‘review\_img\_urls’
* ‘owner\_answer’
* ’ owner\_answer\_timestamp\_datetime\_utc’
* ‘review\_link ‘
* ’ review\_datetime\_utc’

As part of the data reduction the columns 'query', 'google\_id', 'location\_link', 'reviews\_link','reviews\_per\_score', 'review\_id', 'review\_img\_urls','author\_image', 'owner\_answer\_timestamp', 'owner\_answer\_timestamp\_datetime\_utc', ‘review\_datetime\_utc', 'author\_title','review\_img\_url','author\_link','review\_timestamp' were deleted to maintain the privacy of the reviewers in the dataset and to focus on the following 6 categorical variables: ‘name’ ( business name), ‘place\_id’, ‘rating’, ‘review\_id’ , ‘owner\_answer’, ‘review\_text’ and ’ review\_likes’ as part of the NLP processing for this report. The name of the restaurant, the place id and the reviewer id, are the independent variables, while the rating, the owner answer, the number of reviews likes and the actual review text are dependent variables.

The following libraries were imported for various data analysis, preprocessing, and machine learning tasks: NumPy (numeric computations), pandas (data manipulation), seaborn (data visualization), Matplotlib (plotting), NLTK (Natural Language Toolkit for text processing), scikit-learn's TfidfTransformer and CountVectorizer (feature extraction for text data), and train\_test\_split (data splitting for machine learning). Additionally, warnings have been filtered out, and the inline display of Matplotlib plots is enabled. Libraries related to classification reporting and confusion matrix are also included. String processing tools, such as tokenization, stemming, and lemmatization, are available through the NLTK library.

## Exploratory data analysis

Exploratory data analysis was conducted to perform the initial investiagtions on the dataset to discover pattern, trends, and correlation between variables. It was set out in the format of fig. 3 below. The information gained allowed summary statistics and graphical representations of the data to be generated. Visualizations of review length, the distribution of the word count, and the sentiment polarity, stop word distribution, and character count distribution were generated to judge if the dataset is skewed in any way and allow an overview of the data. Both univariate and bivariate analysis were undertaken, to understand each variable individually and how they are related to another.

A diagram of a diagram of a diagram

Description automatically generated

Figure Exploratory data analysis structure

Missing data can distort results and reduce statistical power. Reviews are the main information in this dataset. If any row is missing the review, it was deleted. Estimating or imputation of the missing data would not make sense in this instance. The code *df. isnull().sum()* was used to calculate and display the count of missing values in each column the DataFrame. Other columns with missing data included: ‘review\_img\_url’ and owner\_answer\_timestamp. These were also not required for the dataset analysis and were deleted as part of the data reduction.

**Algorithm 1:** Exploratory Data Analysis, EDA

#Univariate Analysis

1. Describe the dataset, data types, summary statistics, data frame rows and columns

***df.info()***

***df.describe()***

***df.shape()***

1. Create histograms with seaborn of single variables, google hex colour scheme

**sns.distplot(df["rating"], color = '#0F9D58')**

**sns.distplot(df["review\_likes"], color = '#0F4B400)**

**sns.distplot(df["char\_count"], color = '#04285F4)**

1. Create word clouds for reviews based on start rating

**consolidated=' '.join(word for word in df['review\_text'][df['rating']<4].astype(str))**

**wordCloud=WordCloud(width=1600,height=800,random\_state=21,max\_font\_size=110)**

**plt.figure(figsize=(15,10))**

**plt.imshow(wordCloud.generate(consolidated),interpolation='bilinear')**

1. Tabulate tables insert code, most common words

#Bivariate Analysis

1. Create correlation heatmaps

**heatmap = sns.heatmap(df.corr()[['vadar compound']].sort\_values(by='vadar compound', ascending=False), vmin=-1, vmax=1, annot=True, cmap='Blues')**

1. Insert pair plot code

Summary statistics in table 1 below using the .descibe() function provided a high level overview of certain information such as such as count, mean, standard deviation, minimum and maximum values and the quantiles of the reviews themselves, the rating and the review likes. [72].

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **reviews** | **rating** | **author\_id** | **owner\_answer \_timestamp** | **review\_rating** | **review \_timestamp** | **review\_likes** | **reviews\_id** |
| **count** | 1.0E+03 | 1.0E+03 | 1.0E+03 | 1.7E+02 | 1.0E+03 | 1.0E+03 | 1.0E+03 | 1.0E+03 |
| **mean** | 2.7E+03 | 4.4E+00 | 1.1E+20 | 1.7E+09 | 4.4E+00 | 1.7E+09 | 7.7E-01 | 2.9E+17 |
| **std** | 2.5E+03 | 2.4E-01 | 5.4E+18 | 1.4E+07 | 1.1E+00 | 3.2E+07 | 1.5E+00 | 5.1E+18 |
| **min** | 8.5E+02 | 3.9E+00 | 1.0E+20 | 1.6E+09 | 1.0E+00 | 1.5E+09 | 0.0E+00 | -9.1E+18 |
| **25%** | 1.6E+03 | 4.2E+00 | 1.0E+20 | 1.7E+09 | 4.0E+00 | 1.7E+09 | 0.0E+00 | -4.3E+18 |
| **50%** | 2.1E+03 | 4.4E+00 | 1.1E+20 | 1.7E+09 | 5.0E+00 | 1.7E+09 | 0.0E+00 | 7.1E+17 |
| **75%** | 3.0E+03 | 4.5E+00 | 1.1E+20 | 1.7E+09 | 5.0E+00 | 1.7E+09 | 1.0E+00 | 4.0E+18 |
| **max** | 1.3E+04 | 4.8E+00 | 1.2E+20 | 1.7E+09 | 5.0E+00 | 1.7E+09 | 1.7E+01 | 9.0E+18 |

Table 1 Summary statistics of dataset

After removing the missing data, the dataset contained 1328 reviews.

* The minimum rating is 2.8 stars, and the maximum featured rating is 4.9 stars.
* The mean number of review likes is 0.66, (less than 1 per review)

#### Univariate Analysis

For the univariate analysis, charts such as histograms, wordclouds and density plots were used to visualize the data using Matplotlib and Seaborn libraries.

A graph with green lines

Description automatically generated

Figure Histogram of Ratings distribution

Review ratings showed a range of between 3.8 and 5 stars, with a mean value of 4.3 in fig 4.

* The ratings are normally distributed with most of the reviews having a rating of between 4.2. and 4.6.
* This may indicate a high number of falsly positive reviews if the reviews are found to be inauthentic
* It may also indicate that the standard of restaurants in the datast is high and consumers are satisfied with the experience.

The range of ‘review likes’ is spread from 0 to 17.5 in table 2 below, with a skew to the right.

* This shows a low level of engagement among reviewers. Using the .value\_counts () function shows 608, 46% reviews of the total 1328, have 0 likes. 233, 18% have 1 like and 75, 0.06% have 2 likes.

|  |  |
| --- | --- |
| review\_likes | count |
| 0 | 860 |
| 1 | 281 |
| 2 | 95 |
| 3 | 38 |
| 4 | 24 |
| 5 | 12 |
| 6 | 8 |
| 7 | 5 |
| 17 | 2 |
| 13 | 1 |
| 14 | 1 |

Table 2 Count of review likes

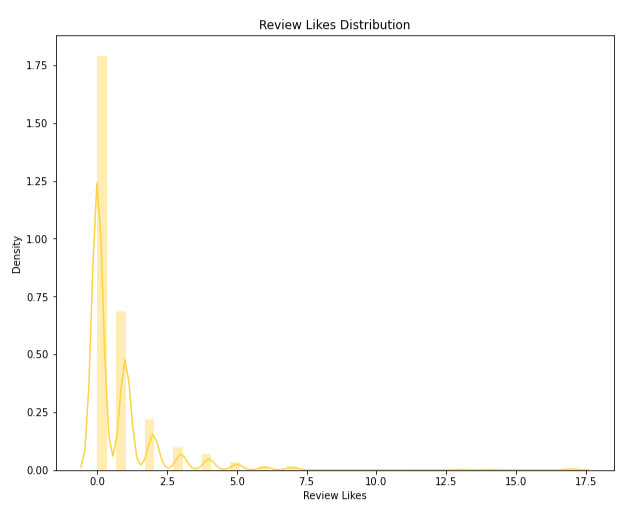


Figure Histogram of review likes distribution

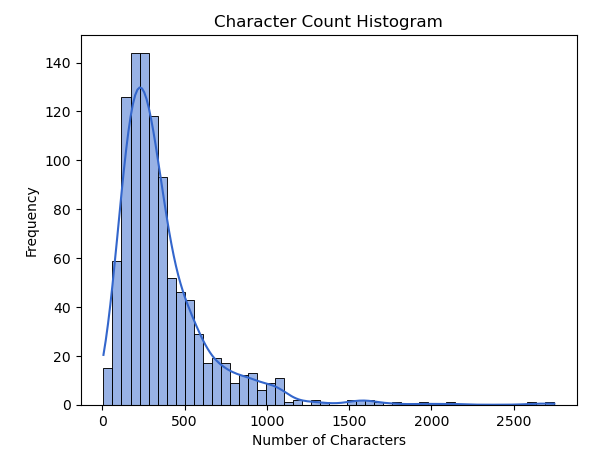


Figure Histogram of character count distribution

Character count per review ranged from 0 to 2500 characters with a distribution skewed postivly to the right in fig. 6. The distribution's tail is skewed to the right, indicating that there are relatively fewer extreme values on the right side and a concentration of lower values on the left side. Over 140 reviews have a character count between 0 and 500. This can indicate either the reviewers are leaving less detail per review, or that the reviews are positive and to the point with no issues highlighted.



Figure Word count of frequent words

From the above word cloud in fig. 7, the following observations were made

* The most frequented words are food, service, staff and Dublin. This suggest that the quality of the food, the service in the restaurant from staff and the location are important to customers.
* A lot of the words in the word cloud have positive connotations such as tasty, friendly, good, quality, excellent, incredible and highly recommend . This suggest most of the reviews are highly rate and have a positive sentiment rating

#### Bivariate Analysis

For the bivariate analysis a combination of correlation heatmaps, pairplots and countplots were used to understand how the variables are related to one another and asses the dependency relationships.

A blue and white graph

Description automatically generated

Figure Heatmap of variable correlation with Sentiment

The above correlation heatmap in cmap colour ‘Blues’, fig 8 shows each variable and it’s correlation with the ‘vadar compound’, (sentiment) variable

* Review rating is the darkest at 0.66 correlation, followed by review likes. This indicates that a positive review tends to get more ratings and likes, than a negative one.

A green bar chart with white text

Description automatically generated

Figure Heatmap of variable correlation with Review likes

Fig. 9 in cmap colour ‘Greens’ theme shows each variable’s level of correlation with ‘Review likes’.

* ‘Check 5‘ (checks if the owner has replied to the review) and ‘check 7‘ (checking number of details) are negativly correlated at -0.0064 and -0.19. This indicates that the polarity of a review has no effect on whether or not the owner of the buisness engages with the review or is related to the number of details they have included in their review.
* Review length and punctuation count are similarly correlated at 0.24 and 0.22 respectivly. This indicates that other reviewers appreciate a longer review with more detail.

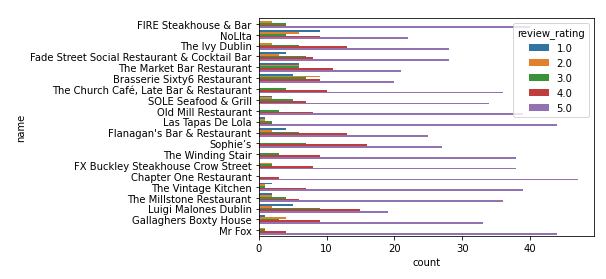


Figure Count plot of restaurants vs. review rating

Fig 10. above shows a seaborn countplot of names of restaurants and the number of stars per restaurant.

* Purple is the most predominant colour showing restaurants in the dataset are highly rated with 5 stars. The Chaper One restaurant has over 40, 5 star reviews
* Sophies restaurant is rated 15 times with 4 star reviews

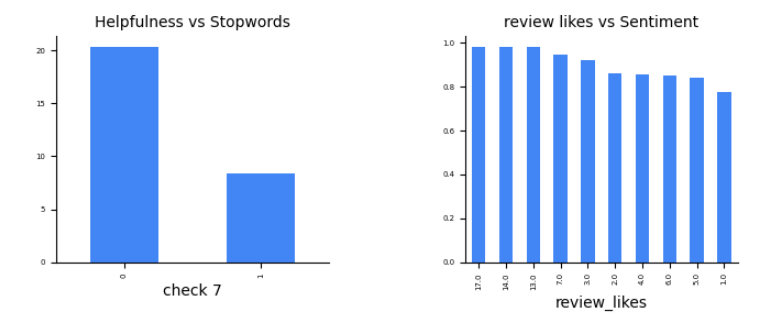


Figure Subplots of helpfulness, stop words, review likes & sentiment

Figure 11 above shows 2 subplots generated with the matplotlib library. The first shows check 7 vs. the number of stop words. Check 7 (review contains at least 10 details, that would be helpful to other customers reading the review). Reviews which meet check 7 (1) have a lower number of stop words, than those that don’t (0). The second subplot shows the number of reviews likes vs the sentiment value. As expected, reviews with a higher number of likes (14 and 17) have a higher positive sentiment rating 0.9 to 1. Reviews with less likes have a slightly lower rating (0.8) which is still very positive.

## Data preprocessing

The text pre-processing was completed as part of the preparation for the application of Natural Language processing, classification, and machine learning. The frequency of punctuation symbols as well as the syntactic and lexical category quantities in each review adds valuable information and contributes to its correct classification. The below flowchart in fig. 12 demonstrates the libraries and methods involved in the preprocessing of the dataset.

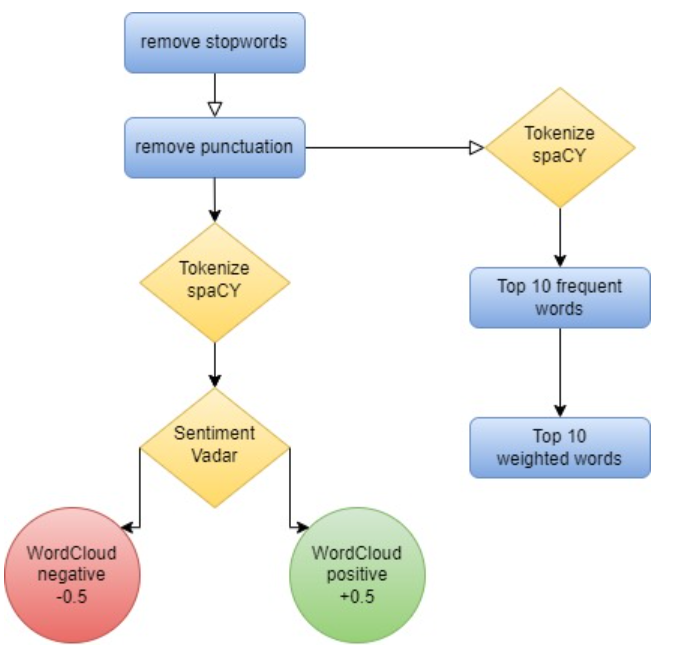


Figure Data Preprocessing structure

The [‘review\_text’] column was also modified to lower case and stop words were removed to allow for more accurate analysis.

Since stop words do not contribute any meaningful information to this text classification model, they were removed from the corpus. This allows for a smaller, more focused dataset. The ‘describe’ function counted 1328 stop words in total in the dataset. This lower-level information also prevents significant words that may contribute to the sentiment analysis later in the analysis from being treated.

The counter class imported from ‘collections’ module was used to process and store punction types and counts for analyzing frequency and generate visualizations of punctuation types. This included exclamation marks, comma’s, dollar signs, hashtags, underscores, emoticons, non-word characters, at symbols, full stops, brackets, and colons. Inauthentic reviews typically have typos, either an excessive amount or complete lack of punctuation in relation to their word count and poor grammar. The punctuation marks of each line were also accumulated in a dictionary and each element was transposed into data frame columns for a greater overview. Three additional columns were added: punctuation count, punctuation list and accumulated punctuation dictionary. This information will be reused later for the rule-based classification.

The Tf-idf library was employed to provide more information on the content of each review. The proportion of occurrences of a certain term to the total number of that term in the dataset provides an insight into emphasis and importance a reviewer has attributed to term. This preprocessing permits a model to learn relationships between words as it is now represented as a vector. Count vectorizer from the sklearn library was used to convert the collection of text into a matrix of tokens. N-gram count of the number of unique words is 976. The count vectorizer was also used to create a bag-of-words representation of the text data which results in a matrix of the reviews. 1328 rows representing the number of reviews and 976 columns representing the unique words. The matrix has non-zero count of 30047 and a relatively low sparsity of 2.32%.

* *sparse matrix shape: (1328, 982)*
* *nonzero count: 30319*
* *sparsity: 2.32%*

The TF-IDF transformed weights were converted to a NumPy array type dataset to present the keywords, weights and sums for each review and demonstrate their context in the complete corpus. The top 10 most frequent words are shown below in table 3. ’Food’ is by far the most common and is featured 1062 times, and ‘good’ is the second with 563 times. Other frequent words featured in the top 10 list have generally positive connotation such as ‘great’, ‘nice’ and ‘delicious.

|  |  |
| --- | --- |
| **Term** | **Occurrences** |
| food | 1062 |
| good | 563 |
| service | 555 |
| great | 529 |
| staff | 439 |
| place | 417 |
| restaurant | 347 |
| nice | 331 |
| would | 284 |
| delicious | 278 |

Table 3 Top 10 frequent words

The top 10 most heavily weighted words are shown below in table 4. ‘Food’ is naturally the heaviest weighted word at 0.0625 and ‘good’ being second at 0.0485. Other heavily weighted words include staff, place and more positive adjectives which provides evidence that the reviews in the dataset are relevant to the subject and overall positive in nature.

|  |  |
| --- | --- |
| **Term** | **Weight** |
| food | 0.062 |
| good | 0.048 |
| great | 0.046 |
| service | 0.042 |
| staff | 0.036 |
| place | 0.035 |
| nice | 0.034 |
| restaurant | 0.029 |
| delicious | 0.029 |
| amazing | 0.027 |

Table 4 Top 10 weighted words

The spaCY package was used to classify the column ‘review\_text’ into named entities and their labels such as locations, cardinal numbers, nouns and dates as seen in fig. 13 below. The displacy visualizer allows a sentence to be broken up and its dependencies to be examined. This package was useful is analyzing the structure of a sentence and to check for specific details such as dates and quantities. Inauthentic reviews will typically lack any specific information that it uniquely relevant to the service and will contain generic comments that suit any restaurant such as ‘Great service’ or ‘Good place for an occasion’. Each named entity such as GPE (location), cardinal, quantity, date and time were stored in an additional column ‘named entities.

The code **‘displacy.render(nlp(str(df['named\_entities'])), jupyter=True, style='ent')’** allows the named entities to be annotated based on their entity type and displayed using and ‘ent’ stype visualization in fig 13 below.

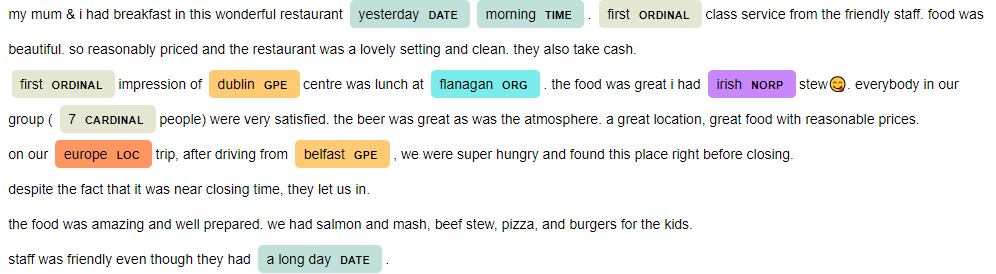


Figure Review with displacy visualizer

The dependency parsing DEP was also generated inside the jupyter notebook. This is particularly useful in semantic role labeling (SRL) and information extraction. The arc label in fig. 14 below describes the type of syntactic relation which connects a dependent child word to a head variable, such as punctuation, meta data, or appositional modifier of a noun.

A diagram of a diagram

Description automatically generated

Figure Named entity structure visualization

Data enrichment was executed to extract more information from the review text column as part of the NLP process.

Sentiment Analysis was completed with Vadar, (Valence Aware Dictionary and sentiment reasoner) and Sentiment Intensity Analyzer from the nltk package. Since this tool is a lexicon and operates on rule-based sentiments, it is particularly suited to social media language, which is appropriate for this dataset. This added 2 additional variables to the dataset; vadar compound which is a numerical variable indicating sentiment between 0 and 1 and a second vadar sentiment is a categorical variable; positive, neutral, and negative depending on the individual review. Thresholds were set to +/-0.5. Through consumer sentiment analysis, companies can detect eh polarity of the review, gage the reaction of competitors and gain insight from their consumers.

A graph with blue squares

Description automatically generated

Figure Sentiment score vs. Rating

Like the other visualization’s, fig 15 shows a single side to the data. The high count of positive reviews with a vadar sentiment score of between 4 and 5 indicate that the reviewers are generally happy with the experience, or they maybe be gushing or overexaggerating their feelings to present a different picture of the restaurant.

|  |  |
| --- | --- |
|  | vadar compound |
| count | 1328 |
| mean | 0.783034 |
| std | 0.360879 |
| min | -0.9493 |
| 25% | 0.7964 |
| 50% | 0.92 |
| 75% | 0.9638 |
| max | 0.9991 |

Table 5 Sentiment summary statistics

Analyzing the vadar sentiment results in table 6, there is 1242 positive reviews, 66 negative reviews and 20 neutral reviews. Vadar compound shows a mean sentiment of 0.78, a min of -0.97 and a max of 0.99 from table 5 below.

|  |  |
| --- | --- |
| vadar sentiment | count |
| positive | 1242 |
| negative | 66 |
| neutral | 20 |

Table 6 Sentiment count



Figure Word cloud of positive reviews

According to the word clouds for positive (above) in fig.16 and negative words(below) in fig. 17 reviews, the words service, friendly, delicious, recommend, atmosphere and staff appeared most frequently among the positive reviews, while ‘staff, rude, poor, disappointed, expensive and unwelcoming’ appeared most frequently in the negative reviews. These indicate that the staff are performing well, are welcoming and the food is appreciated in the restaurants which have highly rated reviews, however there is controversy over the staff manner and the price of the meals.



Figure Word cloud of negative reviews

Rule based classification

8 rule-based classifiers were developed based on the research completed in the literature review using predefined linguistic rules and patterns. These rules aid in categorizing the reviews into specific groups based on the preprocessed data. The extracted categorical and quantitative data will be used to determine if a review is authentic or not, then the result will be compared to machine learning models. None of the rules are individually exhaustive, rather each of them is an indicator that the review may need to be flagged and contains some of the common warning characteristics of an inauthentic review. They are all equally weighted. If a check for inauthenticity succeeds, the review is marked with a binary output of 1, if it fails 0. The sum of the characteristics was chosen as the decisive factor for determining a final pass or fail. The structure of these checks is shown below in fig 18.

A diagram of a check process

Description automatically generated

Figure Rule based classification structure

* ‘Check 1’ examines if the dataset contains multiple reviews for the same restaurant from the same person using if-else statements. Is the number of places reviewed, less than the review count for an individual author id? Leaving multiple reviews is an indicator that the reviewer has not visited the business and either has a financial incentive or a personal agenda against the business.
* ‘Check 2’ examines if the author has submitted more than 1 review in the dataset. Serial reviewers maybe be looking for free gifts from a business or working for a specific platform.
* ‘Check 3’ uses if/else statements to determine whether a reviewer leaves reviews that are extremely positive or negative based on their average vadar compound result from the sentiment analysis. If their average is less that -0.6 or greater than 0.99, all their reviews were flagged. Highly polarized reviews are another red flag, that may indicate the reviewer is biased in their opinion and is not basing the the review on a general experience.
* ‘Check 4’ uses string punctuation to count the number of punctuation marks per review and flags if the count was greater than 10.
* ‘Check 5’ reads the rows of the review to see if an owner has replied to the review. If the owner has acknowledged and engaged with the review, it was taken as a sign that the review was genuine.
* ‘Check 6’ counts the number of characters in a review length. This is a significant point to distinguish spam reviews. If the review substance is excessively short, we can assume the reviewer did not consider the restaurants experience fully. Threshold was set to 150 characters.
* ‘Check 7’ uses the preprocessing completed with the spaCY package to assess the level of detail in each review. That is whether the reviewer has left specific details such as names, locations, dates, times and percentages. Separate functions were written to count each type of detail per row of review. If this count was less than 10, the review was flagged.

The final check takes the sum of the Boolean results of checks 1 to 7 and tests if the integer is greater than 4, that is, that the review has succeeded positively in 4 of the 7 tests at least (more than half). If yes, the review is labelled ‘inauthentic’, otherwise it is labelled ‘authentic’ in a new column ‘label’.

|  |  |
| --- | --- |
| state | count |
| authentic | 1067 |
| inauthentic | 261 |

Table 7 Results rule based classification

The pseudocode for the rule-based classification is shown below.

**Algorithm 2** Rule Based Classification Model

Initialize the check values

check1 = 0

check2 = 0

check3 = 0

check4 = 0

check5 = 0

check6 = 0

check7 = 0

Perform checks and store results

1. If row[‘count\_of\_places\_reviewed’] < [‘review\_count’]:

check 1= 1

else

check 1 = 0

2. if row[‘review\_count’] >1

check 2 =1

else

check 2 =0

3. if row[‘avg\_vadar\_compound’] <= -0.6 or row[‘avg\_vadar\_compound’] >=0.95

check 3 =1

else

check 3=0

4. if row[‘punctuation\_count’] >10

check 4=1

else

check 4=0

5. df[‘check 5’] =np.where(pd.isna(df[‘owner\_answer’]) | (df[‘ownder\_answer’] == ‘’), 1,0)

6. if row[‘char\_count’] <150

check 6=1

else

check 6=0

7. if row[‘noun\_count’] +row[‘date\_count’] +row[‘ordinal\_count’]+row[‘location\_count’]

+row[percent\_count] < 5

check 7=1

else

check 7=0

Initialize count

count = 0

8. def sum\_check(row)

return if row[‘check 1’]+row[‘check 2’] +row[‘check 3’] +row [‘check 4’] +row[‘check 5’] +row[‘check 6’] +row[‘check 7’] >3

9. df[‘label’]=df.apply(sum\_check, axis+1)

# Check sum count and count number of each label, true/ fake

10. df[‘label’]. value\_counts()

## Machine Learning

The dataset was further subset for applying the machine learning to ensure only the relevant columns were included. Columns: ‘Name’, ‘rating’, ‘author\_id’, ‘label’, and ‘review\_text’ were saved as ‘dataset 3’. An additional column called target, which will hold our target variable was created. Inauthentic reviews will be assigned a 1, and authentic reviews will be assigned a 0. The ‘review\_text’ column was split into the training and test datasets with 30% being assigned to the test datast. The model will be trained to predict the target value based on this column alone, identical to the rules-based classification method.

Both supervised and unsupervised machine learning approaches were undertaken.

### Supervised Learning Method

A dictionary of classification models was created, which included XGBClassifer, CatboostClassifier, LinearSVC, MultinomialNB, LGBMClassifier, RandomForestClassifier, DecisionTreeClassifier, ExtraTreeClassifier, AdaBoostClassifier, KNeighborsClassifier, RidgeClassifier, SGDClassifier, BaggingClassifier, BernoulliNB. Cross validation was performed using the different classifiers and their performance was evaluated in terms off the ROC AUC (Receiver Operating Characteristic Area Under the Curve). Since the AUC is a widely used measure of the accuracy of the diagnostic test, it is suitable for this application The high AUC value indicates the binary classifiers are capable of distinguishing between the different classes by measuring the separability. A loop records the ROC AUC score of each classifier including their run time.

A set of hyperparameters were defined as ‘param\_grid’ and ‘GridsearchCV’ from the sklearn package was used to methodically search through the best possible combination in order to improve the ROC AUC score.

The performance of the classifiers was evaluated with metrics such as accuracy, the precision, the recall, and the ROC/AUC score.

### Unsupervised Learning Method

An unsupervised learning algorithm, Kmeans was employed to divide the review data into clusters and perform detailed analysis of the clusters to categorize reviews as fake or real. This method of vector quantization aims to partition n observations into k clusters, where each observation belongs to the cluster with the nearest mean.

Dataset, ‘df1’ was created with the columns: ‘punctuation\_count’, ‘review\_likes’, ‘rating’, ‘avg\_vadar\_compound’, ‘char\_count' and ‘check 7’ (number of details/review) as a subset of the original dataframe, ‘df’. Max clusters limit was set to 7 and the elbow technique was carried out to select the optimal number of clusters. In fig. 19 below, the within-cluster-sum-of-squares (WCSS) values are shown on the y-axis and the number of clusters (k) is shown on the x-axis. The elbow point and optimal value of K in this case was selected at value 4, after which the value of WCSS remains constant to the x-axis.

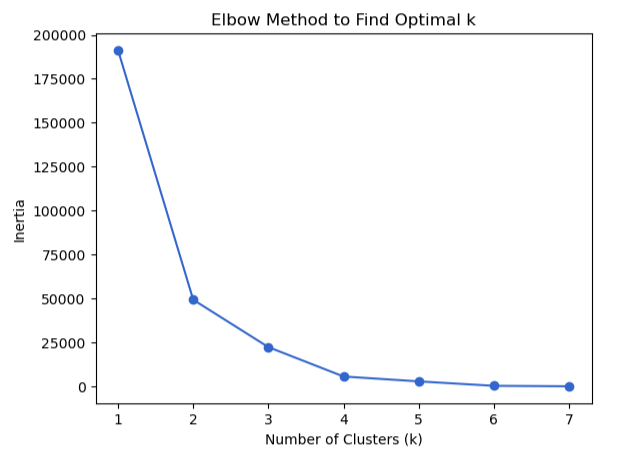


Figure Elbow method for optimal k

The KMeans object was initialized to 4, and the random seed for reproducibility was set to 42. The cluster assignments were added as a new column ‘cluster’ to the dataframe ‘df1’. The cluster centers of each cluster were obtained using the ‘kmeans.cluster\_centers\_’ command, which provides further insight into the characteristics of each cluster. As seen in table 8 below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | punctuation\_ count | review\_likes | rating | avg\_vadar\_ compound | char\_count | check 7 |
| 0 | 5.666667 | 0.666667 | 4.233333 | 0.764273 | 219 | 0.666667 |
| 1 | 13 | 0 | 4.4 | 0.764273 | 584 | 0 |
| 2 | 3 | 0 | 4.1 | 0.764273 | 83.5 | 1 |
| 3 | 11 | 0 | 4.4 | 0.764273 | 376 | 0 |

Table 8 Cluster characteristics

Seaborn was utilized to provide some subplot visualisations of each cluster, to understand their characteristics. Below in fig. 20 is a histogram with KDE showing the distribution of likes across the dataset. The boxplots display the distribution of likes across each cluster.

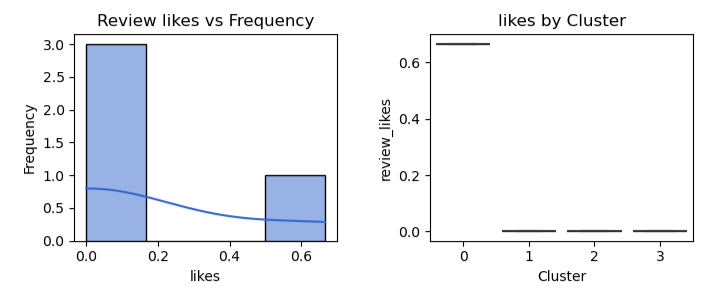


Figure Review likes vs Frequency per cluster

In this case, approximately 70% of the reviews have less than 0.2 likes on average, which 30% have between 0.5 and 0.65 likes. Cluster 0 has more likes compared to other clusters which are quite extreme values. Cluster 0 is largely different from other clusters. A high quantity of likes is another flag for inauthenticity.

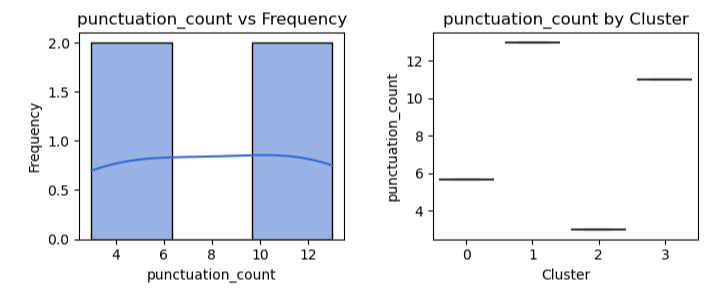


Figure Punctuation count vs Frequency per cluster

The punctuation counts above in fig 21. above is equally split, 50% of the reviews have between 0-6 punctuation counts per review. The other 50% have between 10 and 12 per review. Cluster 0 and cluster 2 have less punctuation compared to 1 and 3. A high punctuation count is another flag for inauthenticity.

|  |  |
| --- | --- |
| cluster | punctuation count |
| 0 | 5.7 |
| 1 | 13 |
| 2 | 3 |
| 3 | 11 |

Table 9 Punctuation count per cluster

Table 9 above show cluster 1 and 3 have punctuation counts of 13 and 11 respectively, while cluster 0 and cluster 2 have counts of 5.7 and 3 respectively.

A comparison of a blue square with a blue square with black text

Description automatically generated with medium confidence

Figure Sentiment vs Frequency per cluster

Fig. 22 above shows no change in sentiment data across the clusters. This indicates that each cluster contains an equal spread of both positive and negative reviews.

A comparison of a graph

Description automatically generated

Figure Level of detail vs Frequency per cluster

Fig. 23 above shows the distribution of column ‘check 7’ across the dataset and per cluster. Check 7 counts the number of named entities such as nouns, dates and place names per review, in other words the number of details. Cluster 0 and 2 contain the highest number of details per review, while cluster 1 and 3 contain lower number of details per review. A lack of details indicates generic comments and is a sign that the reviewer hasn’t physically visited the restaurant.

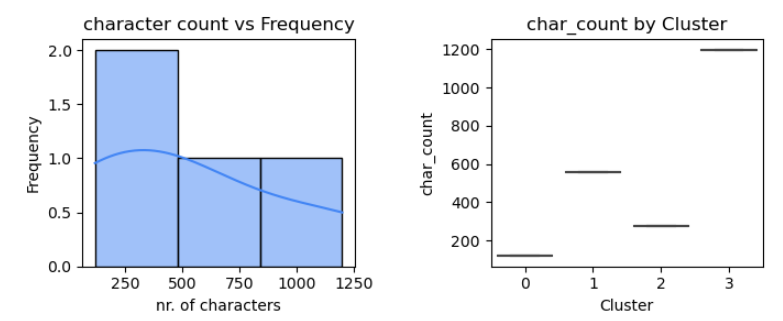


Figure Character count vs Frequency per cluster

The character count above in fig. 24 shows a higher character count in clusters 1 and 3. A long winded, irrelevant review is another flag for inauthenticity.

# Results and Evaluation

## Data pre-processing

Both univariate and bivariate analysis were carried out through the exploratory data analysis. From the univariate analysis, the following can be observed:

* The dataset contains 1328 rows and 26 columns after missing data removed
* Histograms generated show a distribution of ratings from 2.8 to 4.9 stars
* The number of likes per review in the dataset show 46% reviews of the total 1328, have 0 likes. 233, 18% have 1 like and 75, 0.06% have 2 likes
* Character count per review ranged from 0 to 2500 characters. Over 140 reviews have a character count between 0 and 500
* Word clouds from the column ‘review\_text’ showed the most frequented words are food, service, staff and Dublin. This suggest that the quality of the food, the service in the restaurant from staff and the location are important to customers

From the bivariate analysis:

* The variables correlated to a high sentiment review were the star rating (0.66) and the review likes (0.078)
* The level of review likes was positively correlated to the character count (0.24) whether the owner had responded to the review and the level of details (check 7) in the review (0.22). This indicates that other reviewers appreciate a longer review with more detail.
* A countplot of the restaurants vrs. review rating showed most restaurants have very highly rated reviews. . The Chaper One is by far the highest with over 45 5 star reviews. La Tapas and Mr. Fox is the next highest with over 40 5 star reviews

## Rule-based classification

* Check 1, Has the reviewer submitted more than 1 review in dataset?

75 reviewers were flagged for leaving multiple reviews in the dataset. 842 reviewers submitted just 1 review

* Check 2, Has the reviewer submitted multiple reviews for the same place?

3 reviewers were flagged for leaving multiple reviews for the same place, the majority (914) were not flagged.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **author\_id** | **review\_count** | **count\_of\_places\_reviewed** | **check 1** | **check 2** |
| 1.05589E+20 | 3 | 3 | 1 | 0 |
| 1.02056E+20 | 2 | 2 | 1 | 0 |
| 1.15269E+20 | 2 | 2 | 1 | 0 |
| 1.01958E+20 | 2 | 1 | 1 | 1 |
| 1.06966E+20 | 2 | 2 | 1 | 0 |
| 1.15179E+20 | 2 | 2 | 1 | 0 |
| 1.01845E+20 | 2 | 2 | 1 | 0 |
| 1.04228E+20 | 2 | 2 | 1 | 0 |
| 1.04233E+20 | 2 | 2 | 1 | 0 |
| 1.06839E+20 | 2 | 2 | 1 | 0 |

Table 10 Top 10 reviewers ordered by review count

Table 10 shows the top 10 reviewers, ordered by reviewed the review\_count. From table 10, we can see author\_id 1.05589E+20 has submitted the highest number of reviews in the dataset (3), for 3 different places. This means check 1 is flagged but not check 2. Author 1.01958E+20 has submitted 2 reviews for the same place and is flagged in check 1 and 2.

* Check 3, Does the review have an extreme sentiment polarity value?

Of the 1328 reviews, 25 were deemed to have an extreme sentiment value. That is an average vadar compound value of greater than 0.95 or less than -0.06.

* Check 4, does the review have a high punctuation count ( =greater than 10)?

719 reviews have punctuation counts less than 10. 281 reviews have punctuation counts greater than 10 and were flagged for check 4.

|  |  |  |  |
| --- | --- | --- | --- |
| **name** | **author\_id** | **punctuation\_count** | **check 4** |
| Mr Fox | 1.0728E+20 | 79 | 1 |
| NoLIta | 1.0320E+20 | 51 | 1 |
| The Winding Stair | 1.1640E+20 | 51 | 1 |
| Luigi Malones Dublin | 1.1336E+20 | 45 | 1 |
| Old Mill Restaurant | 1.0787E+20 | 43 | 1 |
|  |  |  |  |

Table 11 Top 5 restaurants order by punctuation count

Table 11 is ordered by punctuation count. Author id 1.728E+20 left a review for the Mr. Fox restaurant with the highest punctuation count (79).

* Check 5, Has the business owner verified the review by responding to it?

In 172 reviews in the dataset, the owner has responded to the review and engaged with the reviewer. In the majority of other reviews, (828), the owner has not responded to the review, and the reviews were flagged for check 5.

* Check 6, Is the review less than 150 characters? If the feedback is excessively short, it was taken into consideration that the reviewer did not genuinely experience the restaurant

854 reviews had character counts were greater than 150 characters. 146 reviews were less and were flagged for check 6.

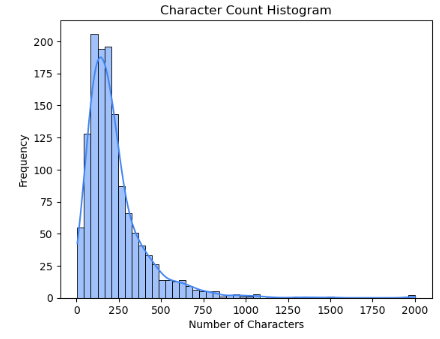


Figure Character count distribution

* Check 7, Is the review helpful to other potential customers by containing detailed information?

Check 7 takes a count of the entities of the review, gained from the spaCY package in the pre-processing section, that is nouns, numbers, dates, locations. If the sum is less than 10 ie. If there were less than 10 details in the review, the review is flagged for check 7. Using the value\_counts function in python showed 738 reviews have more than 10 details, 590 reviews have less than 140 characters.

* Overall, 1067 reviews passed the authenticity check, that it, the review met less than half of the criteria for inauthenticity. 261 reviews met at least 4 of the checks for inauthenticity.

# Machine learning

## Supervised

Due to the size of the dataset and a limited hardware, the dataset was reduced to a smaller sample size to execute the machine learning. 200 random rows were selected to be used for the machine learning.

A machine learning pipeline was used, consisting of two main components: TF-IDF vectorizer for text feature extraction on the ‘review\_text’ column and SGD classifier for classification on the authenticity of the review. The pipeline was fitted to the training data (X\_train, y\_train) to make predictions on the test data (X\_test).

The set of hyperparameters ‘param\_grid’ were defined for the grid search CV. Learning rate range was specified between 0.1 and 0.01 based on the weight at the end of each batch. The number of estimators was set between 50 and 300 decision trees. Max depth for the gradient boosting was evaluated at odd values between 1 and 7 (1,3,5,7), and the subsample range was chosen between 0.6 and 1.2. The hyperparameter were chosen based on an optimum ROC output value.

The grid search outputted the best parameters as below in Table 12. The best ROC score was 0.6333, which is a poor result. A value of 0.63 means it performs better than a random example only 63% of the time. The value is relatively close to the upper left corner of the graph where specificity and sensitivity are close to 1.

|  |  |
| --- | --- |
| **Parameters** | **Value** |
| learning rate | 0.1 |
| n estimators | 50 |
| max depth | 5 |
| subsample | 0.6 |

Table 12 Supervised ML parameters

|  |  |  |  |
| --- | --- | --- | --- |
| **model** | **run\_time** | **roc\_auc** | **roc\_auc\_std** |
| MultinomialNB | 0 | 0.706845 | 0.028274 |
| SGDClassifier | 0 | 0.696429 | 0.14881 |
| LinearSVC | 0 | 0.65625 | 0.049107 |
| AdaBoostClassifier | 0.01 | 0.643601 | 0.008185 |
| RidgeClassifier | 0 | 0.642857 | 0.047619 |
| RandomForestClassifier | 0.01 | 0.62872 | 0.042411 |
| DecisionTreeClassifier | 0 | 0.616071 | 0.03869 |
| CatBoostClassifier | 0.29 | 0.592262 | 0.056548 |
| BaggingClassifier | 0 | 0.590774 | 0.004464 |
| XGBClassifier | 0 | 0.555804 | 0.073661 |
| ExtraTreeClassifier | 0 | 0.528274 | 0.052083 |
| LGBMClassifier | 0.01 | 0.523065 | 0.02753 |
| KNeighborsClassifier | 0.01 | 0.504464 | 0.081845 |
| BernoulliNB | 0 | 0.342262 | 0.104167 |

Table 13 Model results, Supervised ML

Table 13 shows a list of the 14 models ranked by ROC AUC score. MultinomialNB was ranked highest with a score of 0.71 followed closely by SGDClassifier with a score of 0.696. All of the results are generally quite poor to fair in the range of 0 to 0.7. Further hyperparameter tuning might improve a model result above 0.8. The final model accuracy was estimated at 0.87, with a precision of value 0, a recall of value 0.0 and an ROC AUC value of 0.5.

Accuracy: 0.8666666666666667

Precision: 0.0

Recall: 0.0

ROC/AUC: 0.5

## Unsupervised

On the basis of analysis from pipeline of k means clustering,

* Clusters 1 and 3 have high punctuation counts
* Cluster 0 has a higher number of likes compare to cluster 1,2 and 3
* Clusters 1 and 3 have low level of detail
* Cluster 0 and 2 have a low character count
* Cluster 3 has a high character count but low level of detail
* Sentiment and star rating is equal across the board

Clusters 1 have 3 have more flags for inauthenticity than clusters 0 and 4 overall.

# Conclusion

The paper discusses how sentiment analysis facilitates online buying power, aided by machine learning techniques. Sentiment analysis enables businesses to understand customers’ aggregate opinions and attitudes towards certain products by distinguishing the polarity of the reviews, and helps customers make the correct and informed decision by supporting their research. according to the word cloud for positive and negative words in reviews. The paper gives some clear suggestions for tackling with the most frequently appeared complaints - the staff manner and the price of the meals based on data submitted by real customers.

1. To generate several machine learning models on the trained dataset and compare their performance results and overall effectiveness with the rule-based classification system

The data was sourced from a combination of google reviews for restaurants in Galway and Dublin Ireland between February 2016 and April 2023. The first research objectives of this thesis was to process and classify online reviews using NLP techniques. The exploratory data analysis shows a large and varied dataset with ratings between 2.8 to 4.0 stars and and sentiment polarity scores of -0.9 and +0.99. This indicates a range of satisfied and unsatisfied customers. Different types of restaurants are included ranging from upmarket steakhouses such as Fire Steakhouse & Bar to pub gastronomy such as Flanagan’s Restaurant & Bar. The Ivy in Dublin is a stylish brasserie while the Fox is a contemporary irish restaurant with a French influence offering an attractive vegetarian menu. The Quay Street kitchen is a small bustling restaurant serving up a range of cuisines in the heart of Galway city.

An in-depth literature review identified potential features which can help tell an authentic review from an inauthentic one such as the review length, the sentiment score, the helpfulness of the review, verification, coherence and readability. These features were taken into account for the rule-based classification design.

There are many types of inauthentic reviews, they range from computer generated to human generated and can be either positive or negative depending on the creator’s intention. The intention of this report is solely to identify methods to improve the accuracy of categorizing these reviews into 1 of 2 boxes; authentic or inauthentic. Several different analytical methods have been employed to accomplish this aim such as sentiment analysis, rule-based classification and machine learning.

This thesis is primarily focused on the actual review text column of the dataset. Further analysis extracted more quantitative information from this column such as the character and punctuation count, the sentiment polarity, the helpfulness score of the review. Other factors such as the number of likes a review had and its star rating were documented and visualized but not taken further.

The rule-based classification system was implemented as part of the research objectives to detect if online reviews are completed by authentic patrons. This approach was judged to be the more successful approach however since this was completed with unlabelled data (regarding whether or not the review is authentic or not), there is no performance statistic to compare the accuracy. Overall, 1067 reviews passed the authenticity check, that it, the review met less than half of the criteria for inauthenticity. 261 reviews met at least 4 of the checks for inauthenticity. These rules can easily be adjusted to be made stricter, such as tightening the acceptable sentiment limits or reducing the limits for character count and punctuation. Additional rules can also be added easily to test for other characteristics of inauthentic reviews.

Machine learning models on the trained dataset were implemented as per the third research object and their performance results and overall effectiveness was compared with the rule-based classification. The supervised machine learning method returned a final accuracy of 0.87 which is a good result for predicting the authenticity of a review against the rule-based classification method. The unsupervised method pointed towards clusters which could then be broken down further by identify author id’s and used to identify problem areas.

Some of the limitation of this thesis include the sample size of the machine learning due to hardware requirements and time limitations to complete this project. Future improvements could involve expanding the rule-based classification such as testing for sarcasm or going into more detail on the level of helpfulness in a review. The algorithms could also be tested on unknown datasets or compare the performance against a labelled dataset. More machine learning could also be undertaken with different models and further hyperparameter tuning such as F1 scores and confusion matrices.

# References

1. Broida, R. (2019). *How to spot fake Amazon reviews*. [online] CNET. Available at: <https://www.cnet.com/deals/spot-fake-reviews-amazon-best-buy-walmart/>.
2. Khalifah (n.d.). *Fakespot | Analyze and identify fake reviews and counterfeits*. [online] www.fakespot.com. Available at: <https://www.fakespot.com/>.
3. Federal Trade Commission (2021). *FTC Puts Hundreds of Businesses on Notice about Fake Reviews and Other Misleading Endorsements*. [online] Federal Trade Commission. Available at: <https://www.ftc.gov/news-events/news/press-releases/2021/10/ftc-puts-hundreds-businesses-notice-about-fake-reviews-other-misleading-endorsements>.
4. Gain, V. (2022). *Amazon sues 10,000 Facebook groups over fake reviews*. [online] Silicon Republic. Available at: <https://www.siliconrepublic.com/business/amazon-fake-reviews-facebook-groups>.
5. OAK, R. (2022). Inside the Underground Market for Fake Amazon Reviews. *Wired*. [online] 2 Nov. Available at: <https://www.wired.com/story/fake-amazon-reviews-underground-market/>.
6. European Consumer Centre Ireland. Available at: <https://www.eccireland.ie/fake-reviews-and-false-ratings-online-shopping/#:~:text=Some%20fake%20reviews%20are%20harder>.
7. European Consumer Centre Ireland. Available at: <https://www.eccireland.ie/fake-reviews-and-false-ratings-online-shopping/#:~:text=Some%20fake%20reviews%20are%20harder>.
8. Azimi, S., Krasnikov, A. and Chan, K. (2023). *Why we usually can’t tell when a review is fake*. [online] npr.org. Available at: <https://www.npr.org/sections/money/2023/03/07/1160721021/why-we-usually-cant-tell-when-a-review-is-fake#:~:text=Azimi%20says%20study%20participants%20tended,the%20readability%20of%20the%20text.>.
9. Dean, K. (n.d.). *Fake Review Watch*. [online] Fake Review Watch. Available at: <https://fakereviewwatch.com/>.
10. Liu, J. (2023). *Method to Facilitate E-Commerce Buying Power by Using Machine Learning Techniques*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/370709857_Method_to_Facilitate_E-Commerce_Buying_Power_by_Using_Machine_Learning_Techniques>.
11. Plotkina, D., Munzel, A. and Pallud, J. (2020). Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research*, 109, pp.511–523. doi:<https://doi.org/10.1016/j.jbusres.2018.12.009>.
12. Amazon (2019). *ReviewMeta.com - Amazon Review Checker*. [online] Reviewmeta.com. Available at: <https://reviewmeta.com/>.
13. Sri, P., Reddy, R. and Analytics (2021). A Fake Review Detection System Using NLP and Machine Learning Techniques. *International Journal of Scientific & Engineering Research*, [online] 12(8). Available at: <https://www.ijser.org/researchpaper/A-Fake-Review-Detection-System-Using-NLP-and-Machine-Learning-Techniques.pdf>.
14. Thubron, R. (2016). *Amazon has removed hundreds of thousand of incentivized reviews since it banned the practice*. [online] TechSpot. Available at: <https://www.techspot.com/news/67171-amazon-has-removed-hundreds-thousand-incentivized-reviews-since.html>.
15. MCCLUSKEY, M. (2022). *Inside the War on Fake Reviews*. [online] Time magazine. Available at: <https://time.com/6192933/fake-reviews-regulation/>.
16. Amazon (2023). *Amazon continues to take action against fake review brokers*. [online] US About Amazon. Available at: <https://www.aboutamazon.com/news/policy-news-views/amazon-continues-to-take-action-against-fake-review-brokers>.
17. Labhishetty, S., Zhai, C., Xie, M., Gong, L., Sharnagat, R. and Chembolu, S. (2022). Differential Query Semantic Analysis. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. doi:<https://doi.org/10.1145/3488560.3498503>.
18. Choi, W., Nam, K. and Yang, S. (2023). *Fake review identification and utility evaluation model using machine learning*. [online] Frontiers. Available at: <https://www.frontiersin.org/articles/10.3389/frai.2022.1064371/full> 2022, Volume 5 .
19. Choudhary, N. (2021). LDC-IL: The Indian repository of resources for language technology. *Language Resources and Evaluation*. doi:<https://doi.org/10.1007/s10579-020-09523-3>.
20. Collobert, R., Weston, J., Com, J., Karlen, M., Kavukcuoglu, K. and Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, [online] 12, pp.2493–2537. Available at: <https://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf>.
21. Davis, E. and Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9), pp.92–103. doi:<https://doi.org/10.1145/2701413>.
22. Fan, Y., Tian, F., Xia, Y., Qin, T., Li, X.-Y. and Liu, T.-Y. (2020). Searching Better Architectures for Neural Machine Translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, pp.1574–1585. doi:<https://doi.org/10.1109/taslp.2020.2995270>.
23. Fang, X. and Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, 2(1). doi:<https://doi.org/10.1186/s40537-015-0015-2>.
24. Fattah, M.A. and Ren, F. (2009). GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *Computer Speech & Language*, 23(1), pp.126–144. doi:<https://doi.org/10.1016/j.csl.2008.04.002>.
25. Gao, T., Dontcheva, M., Adar, E., Liu, Z. and Karahalios, K.G. (2015). DataTone. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. doi:<https://doi.org/10.1145/2807442.2807478>.
26. Mani , I. and Maybury, M.T. (1999). *Advances in Automatic Text Summarization*. [online] MIT Press. Available at: <https://mitpress.mit.edu/9780262133593/advances-in-automatic-text-summarization/>
27. Khurana, D., Koli, A., Khatter, K. and Singh, S. (2022). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82, pp.3713–3744. doi:<https://doi.org/10.1007/s11042-022-13428-4>.

[Accessed 4 Sep. 2023].

1. Marciano, J. (2021). *Fake online reviews cost $152 billion a year. Here’s how e-commerce sites can stop them*. [online] World Economic Forum. Available at: <https://www.weforum.org/agenda/2021/08/fake-online-reviews-are-a-152-billion-problem-heres-how-to-silence-them/>.
2. McKeown, K. (1985). *Text Generation*. [online] *Cambridge University Press*. Cambridge: Cambridge University Press. Available at: <https://www.cambridge.org/core/books/text-generation/EC9075AB5F64A7AFDC5A9513237505CA> [Accessed 14 May 2023].
3. Newatia, R. (2019). *How to Implement CNN for NLP tasks like Sentence Classification*. [online] Medium. Available at: <https://medium.com/saarthi-ai/sentence-classification-using-convolutional-neural-networks-ddad72c7048c>.
4. Ogallo, W. and Kanter, A.S. (2017). Using Natural Language Processing and Network Analysis to Develop a Conceptual Framework for Medication Therapy Management Research. *AMIA Annual Symposium Proceedings*, [online] 2016, pp.984–993. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333323/> [Accessed 4 Sep. 2023].
5. Rehm, G., Sasaki, F., Stein, D. and Witt, A. (1967). *Language technologies for a multilingual Europe TC3 III Translation and Multilingual Natural Language Processing 5. [online] Available at:*[*https://library.oapen.org/bitstream/handle/20.500.12657/28285/1001677.pdf?sequence=1#page=13*](https://library.oapen.org/bitstream/handle/20.500.12657/28285/1001677.pdf?sequence=1#page=13)
6. Gokce, E. (2020). *Sentiment Analysis on Amazon Reviews*. [online] Medium. Available at: <https://towardsdatascience.com/sentiment-analysis-on-amazon-reviews-45cd169447ac>.
7. Liaqat, M.I., Awais Hassan, M., Shoaib, M., Khurshid, S.K. and Shamseldin, M.A. (2022). Sentiment analysis techniques, challenges, and opportunities: Urdu language-based analytical study. *PeerJ Computer Science*, 8, p.e1032. doi:<https://doi.org/10.7717/peerj-cs.1032>.
8. MIAP, C.T.Bs.H. (2019). *Recurrent Neural Networks and Natural Language Processing.* [online] Medium. Available at: <https://towardsdatascience.com/recurrent-neural-networks-and-natural-language-processing-73af640c2aa1>.
9. Wang, D., Zhu, S., Li, T., Chi, Y. and Gong, Y. (2011). Integrating Document Clustering and Multidocument Summarization. *ACM Transactions on Knowledge Discovery from Data*, 5(3), pp.1–26. doi:<https://doi.org/10.1145/1993077.1993078>.
10. Woods, W.A. (1977). *Semantics and Quantification in Natural Language Question Answering*. [online] apps.dtic.mil. Available at: <https://apps.dtic.mil/sti/citations/ADA047600> [Accessed 4 Sep. 2023].
11. OAK, R. (2022). Inside the Underground Market for Fake Amazon Reviews. *Wired*. [online] 2 Nov. Available at: <https://www.wired.com/story/fake-amazon-reviews-underground-market/>.
12. Sri, P., Reddy, R. and Analytics (2021). A Fake Review Detection System Using NLP and Machine Learning Techniques. *International Journal of Scientific & Engineering Research*, [online] 12(8). Available at: <https://www.ijser.org/researchpaper/A-Fake-Review-Detection-System-Using-NLP-and-Machine-Learning-Techniques.pdf>.
13. Riedhammer, K., Favre, B. and Hakkani-Tür, D. (2010). Long story short – Global unsupervised models for keyphrase based meeting summarization. *Speech Communication*, 52(10), pp.801–815. doi:<https://doi.org/10.1016/j.specom.2010.06.002>.
14. Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C.D. and Stamatopoulos, P. (2003). A Memory-Based Approach to Anti-Spam Filtering for Mailing Lists. *Information Retrieval*, 6(1), pp.49–73. doi:<https://doi.org/10.1023/a:1022948414856>.
15. Small, S., Cottrell, G. and M Tanenhaus (1987). *Lexical ambiguity resolution*. Morgan Kaufman Publishers, Inc.,Los Altos, Ca.
16. Y. Li, L. and Qin, B. (2018). *Survey on Fake Review Detection Research*. [online] Available at: <https://www.researchgate.net/publication/327988694_Survey_on_Fake_Review_Detection_Research>.
17. Chao, J. and Zhao, C. (2022). Network Embedding-Based Approach for Detecting Collusive Spamming Groups on E-Commerce Platforms. *Researchgate.net*.
18. Zhang, F., Yuan, S., Zhang, P., Chao, J. and Yu, H. (2022). Detecting review spammer groups based on generative adversarial networks. *Information Sciences*, [online] 606, pp.819–836. doi:<https://doi.org/10.1016/j.ins.2022.05.086>.
19. Barbado, R. and Araque, O. (2019). *A framework for fake review detection in online consumer electronics retailers*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/332110221_A_framework_for_fake_review_detection_in_online_consumer_electronics_retailers>.
20. Lin, K.-C., Miao, W. and Sun, W. (2023). *How Could Consumers’ Online Review Help Improve Product Design Strategy?* [online] Rsearchgate. Available at: [https://www.researchgate.net/publication/372860419\_How\_Could\_Consumers'\_Online\_Review\_Help\_Improve\_Product\_Design\_Strategy](https://www.researchgate.net/publication/372860419_How_Could_Consumers%27_Online_Review_Help_Improve_Product_Design_Strategy)
21. Duma, R.A., Niu, Z. and Nyamawe, A. (2023). *DHMFRD – TER: a deep hybrid model for fake review detection incorporating review texts, emotions, and ratings*. [online] Rsearchgate. Available at: <https://www.researchgate.net/publication/371079473_DHMFRD_-_TER_a_deep_hybrid_model_for_fake_review_detection_incorporating_review_texts_emotions_and_ratings>.
22. Mewada, A. and DEWANG, R.K. (2021). *Research on False Review Detection Methods: A state-of-the-art review*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/353724379_Research_on_False_Review_Detection_Methods_A_state-of-the-art_review>
23. Gryka, P. and Janicki, A. (2023). Detecting Fake Reviews in Google Maps—A Case Study. *Applied Sciences*, [online] 13(10), p.6331. doi:<https://doi.org/10.3390/app13106331>.
24. Cai, M., Du, Y. and Tan, Y. (2023). Aspect‑based classiﬁcation method for review spam detection. *Springer*. doi:<https://link.springer.com/article/10.1007/s11042-023-16293-x>.
25. Saastamoinen, M. and Kumpulainen, S. (2014). *Expected and materialised information source use by municipalofficials: intertwining with task complexity*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/279036301_Expected_and_materialised_information_source_use_by_municipal_officials_Intertwining_with_task_complexity>.
26. Mewada, A. (2022). *A comprehensive survey of various methods in opinion spam detection*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/363282201_A_comprehensive_survey_of_various_methods_in_opinion_spam_detection>.
27. Elzeheiry, S. and Gab Allah, W.A.M. (2023). Sentiment Analysis for E-commerce Product Reviews: Current Trends and Future Directions. *Researchgate*. doi:<https://www.researchgate.net/publication/370986246_Sentiment_Analysis_for_E-commerce_Product_Reviews_Current_Trends_and_Future_Directions>.
28. Liu, J. (2023). *Method to Facilitate E-Commerce Buying Power by Using Machine Learning Techniques*. [online] Available at: <https://www.researchgate.net/publication/370709857_Method_to_Facilitate_E-Commerce_Buying_Power_by_Using_Machine_Learning_Techniques>.
29. Li, L., Zheng, H., Chen, D. and Zhu, B. (2022). Whose reviews are most valuable for predicting the default risk of peer-to-peer lending platforms? Evidence from China. *Electronic Commerce Research*. doi:<https://doi.org/10.1007/s10660-022-09571-7>.
30. Mukherjee, A., Venkataraman, V., Liu, B. and Glance, N. (n.d.). *Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews*. [online] Available at: <https://www2.cs.uh.edu/~arjun/papers/UIC-CS-TR-yelp-spam.pdf>.
31. Zhang, D., Li, W., Niu, B. and Wu, C. (2023). A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information. *Decision Support Systems*, 166, p.113911. doi:<https://doi.org/10.1016/j.dss.2022.113911>.
32. Salminen, J. (2021). Fake Reviews Dataset. *osf.io*. [online] Available at: <https://osf.io/tyue9/>.
33. Hammad, A. and El-Halees, A. (2013). *An Approach for Detecting Spam in Arabic Opinion Reviews*. [online] Available at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b350cd14f88e5848392b6417c731832679472eb8>.
34. Wu, G., Greene, D., Smyth, B. and Cunningham, P. (n.d.). *Distortion as a Validation Criterion in the Identification of Suspicious Reviews*. [online] Available at: <https://snap.stanford.edu/soma2010/papers/soma2010_2.pdf> [Accessed 4 Sep. 2023].
35. Romanov, D., Molokanov, V., Kazantsev, N. and Jha, A.K. (2022). Removing order effects from human-classified datasets: A machine learning method to improve decision making systems. *Decision Support Systems*, p.113891. doi:<https://doi.org/10.1016/j.dss.2022.113891>.
36. Utz, S., Kerkhof, P. and van den Bos, J. (2012). Consumers rule: How consumer reviews influence perceived trustworthiness of online stores. *Electronic Commerce Research and Applications*, 11(1), pp.49–58. doi:<https://doi.org/10.1016/j.elerap.2011.07.010>.
37. Utz, S., Kerkhof, P. and van den Bos, J. (2012). Consumers rule: How consumer reviews influence perceived trustworthiness of online stores. *Electronic Commerce Research and Applications*, 11(1), pp.49–58. doi:<https://doi.org/10.1016/j.elerap.2011.07.010>.
38. Park, D.-H. and Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, 7(4), pp.399–410. doi:<https://doi.org/10.1016/j.elerap.2007.12.001>.
39. Harrison-Walker, L.J. and Jiang, Y. (2023). Suspicion of online product reviews as fake: Cues and consequences. *Journal of Business Research*, 160, p.113780. doi:<https://doi.org/10.1016/j.jbusres.2023.113780>.
40. Banerjee, S. and Chua, A.Y.K. (2023). Understanding online fake review production strategies. *Journal of Business Research*, 156, p.113534. doi:<https://doi.org/10.1016/j.jbusres.2022.113534>.
41. Wang, Y., Zamudio, C. and Jewell, R.D. (2023). The more they know: Using transparent online communication to combat fake online reviews. *Business Horizons*. doi:<https://doi.org/10.1016/j.bushor.2023.03.004>.
42. Costa Filho, M., Nogueira Rafael, D., Salmonson Guimarães Barros, L. and Mesquita, E. (2023). Mind the fake reviews! Protecting consumers from deception through persuasion knowledge acquisition. *Journal of Business Research*, 156, p.113538. doi:<https://doi.org/10.1016/j.jbusres.2022.113538>.
43. Costa, A., Guerreiro, J., Moro, S. and Henriques, R. (2019). Unfolding the characteristics of incentivized online reviews. *Journal of Retailing and Consumer Services*, 47, pp.272–281. doi:<https://doi.org/10.1016/j.jretconser.2018.12.006>.
44. aman.ai. (n.d.). Aman’s AI Journal • CS224n: Natural Language Processing with Deep Learning. [online] Available at: https://aman.ai/cs224n/ [Accessed 5 Sep. 2023].
45. Mahadevan, M. (2022). Step-by-Step Exploratory Data Analysis (EDA) using Python -. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/>.

# Appendices

## Appendix 1-Interview

Interview 1

21.8.23, Bar and Restaurant Manager Galway, Rúibín Bar & Restaurant

Thesis author: Teresa Quain

Several interviews were conducted with people in the service industry. One in particular was from the Bar & Restaurant manager in the Rúibín Bar & Restaurant in Galway city who run a lunch and dinner service for approximately 80 to 90 covers per day.

He runs the front of house and is responsible for interacting with the customers and oversees bookings, food service and the potential customer reviews at the end.

He confirmed that the restaurant does not solicit reviews from a customer. Most reviews for their restaurant are posted on google or TripAdvisor. When asked what percentage of people leave a review, he said it can go either way. Sometimes the customer had a great experience and are satisfied with the meal and will verbalise it to the staff when they are there but will not end up leaving a review. Other times, a different customer will have had the same experience, same menu and will not be happy with the overall service. They may then leave a very negative review. This may then cause overall average ratings for a business to be skewed. He confirmed that they have some reviews which he believes are inauthentic. The reasons for this point of view where he mentioned the reviews contained very generic information or impossible scenarios. While these are logical assumptions there is no way to trace back a reviewer without account access and verify if they have visited the restaurant. His complaints about the review platforms included that reviewers do not require any verification method and remain anonymous. Attempts to contact TripAdvisor to address unjust reviews were not dealt with satisfactorily from the businesses point of view.

This interview confirms the research conducted in the literature review that inauthentic reviews are causing difficulties for businesses. The lack of feedback from TripAdvisor platform does not inspire trust in the software from a business perspective and seems to be more geared towards the number of reviews rather than their quality.

Appendix 2 Participant Consent form

Enhancing the Accuracy of Inauthentic Review Detection using Machine Learning and Sentiment Analysis

Consent to take part in research

I voluntarily agree to participate in this research study.

I understand that even if I agree to participate now, I can withdraw at any time or refuse to answer any question without any consequences of any kind.

I have had the purpose and nature of the study explained to me in writing and I have had the opportunity to ask questions about the study.

I understand that I will not benefit directly from participating in this research.

I agree to my interview being audio-recorded and the transcript shared in the academic research study

I understand that all information I provide for this study will be treated confidentially.

I understand that in any report on the results of this research my identity will be shared with the CCT college and the interviewer only. Personal data will not be shared elsewhere. The identity of people I speak about will not be shared

I understand that signed consent forms and original audio recordings will be retained in [*specify location, security arrangements and who has access to data*] until [*specific relevant period – for students this will be until the exam board confirms the results of their dissertation*].

I understand that a transcript of my interview in which all identifying information has been removed will be retained for [*specific relevant period – for students this will be two years from the date of the exam board*].

I understand that under freedom of information legalisation I am entitled to access the information I have provided at any time while it is in storage as specified above.

I understand that I am free to contact any of the people involved in the research to seek further clarification and information.

Names, degrees, affiliations and contact details of researchers (and academic supervisors when relevant).

*Signature of research participant*

Richard Ruibin, 21 Aug 2023

Signature of participant Date

*Signature of researcher*

I believe the participant is giving informed consent to participate in this study

Teresa Quain, 21 August 2023

Signature of researcher Date

## Appendix 3 Github Version Control

Account: [teresaq1](https://github.com/teresaq1/thesis_Capstone-project_Msc_FakeReviews)

Repository: [thesis\_Capstone-project\_Msc\_FakeReviews](https://github.com/teresaq1/thesis_Capstone-project_Msc_FakeReviews)

Source: <https://github.com/teresaq1/thesis_Capstone-project_Msc_FakeReviews>

