It's Time to Get Real: Investigating the Effectiveness of Linguistic and Sentiment Analysis Features in Identifying Fake Reviews

A Linguistic and Sentiment Analysis of an Amazon Fake Review Dataset researching to what extent a machine learning algorithm can detect fake reviews using only linguistic and sentiment analysis features and which linguistic and sentiment analysis features are most effective.

J.G.C. van Lier

SNR: 2087156

August 11, 2023

https://github.com/JessevanLier/feature_based_fake_review_detection.git

Master Thesis Marketing Analytics

Marketing Department

Tilburg School of Economics and Management

Tilburg University

Master Thesis Supervisors:	dr. Hannes Datta	Tilburg University
	dr. Shrabastee Banarjee	Co-reader Tilburg University

Management Summary

Consumers have used electronic word of mouth extensively over the past few decades on various online platforms to acquire product feedback from prior customer experiences. The scale and detail of online information exchange is unprecedented. However, increasing numbers of customers are dissatisfied and suspicious of digital communications such as online reviews. Fake reviews are a well-recognized problem by businesses and the scientific community. However, despite much interest, fake reviews have proven difficult due to the various review writers, employees, fake-review marketplaces, machines, and customers. Because of this ongoing problem, there is an urge for more transparency and information on detecting fake reviews and their relevant features. Previous studies have shown that linguistic and sentiment analysis features can be used to distinguish between fake and real reviews with great accuracy. To distinguish between fake and real online reviews, this thesis will focus on analyzing textual features like sentiment and linguistics. Furthermore, we will investigate to what extent a machine learning algorithm can detect fake reviews using only linguistic and sentimental features in e-commerce and which linguistic and sentiment analysis features are most effective. This research investigates how textual features can predict fake reviews without the actual text. Although research on the performance of these linguistic features in the context of textual data to detect fake reviews has been done, the effectiveness of lexicons has yet to be explored.

Because fake reviews significantly impact the success of e-commerce retailers, we used an Amazon Dataset for this research. We incorporated the VADER and SentiWordNet lexicons to extract the sentiment analysis features per review and used Natural Language Processing (NLP) to extract the linguistic features. To classify the reviews, we will use the Support Vector Machine (SVM) for the main test and Naïve Bayes (NB) to compare its performances. We created three features' sets using the two lexicons and combining the sets into one compete set. The feature sets include features from the following categories: quantity, complexity, diversity, and sentiment. We conclude that the SVM algorithm has a better performance than NB. The highest accuracy-score of 60% was realized by using 33 features from the combined lexicon list. Furthermore, the presented lists of features for both lexicons and combined that provide insights into the linguistic and sentiment analysis features that are most relevant for detecting fake reviews and enables future research to development models for fake review identification that are more effective and accurate. In short, thanks to our research, there are three feature sets available for researchers to investigate fake reviews further and help consumers and businesses understand the detection of these reviews better to ensure no harm is done in the future.

Preface

Dear Reader,

The result is finally here after a long period of blood, sweat, and tears. First, I would like to express my deepest appreciation to my supervisor Hannes Datta for always asking the right questions, for your expertise in marketing analytics, and the feedback. I know I can be a little chaotic, but thanks to the thesis meetings in the early morning, I always knew what to do the rest of the week.

Also, I would like to extend my sincere thanks to my family and friend for their unconditional support. The long days on the university would have been a lot harder if it wasn't for you.

Wishing you an enjoyable and enlightening reading experience with this thesis.

Jesse van Lier,

Nijmegen, August 11, 2023

Table of contents

Chapter 1: Introduction

§ 1.1. Introduction

"If you do build a great experience, customers tell each other about that. Word of mouth is very powerful" (Bezos, J., 1997). Only a few years after starting Amazon, Jeff Bezos recognized the power of word-of-mouth (WOM) for online retail businesses and implemented this into his business strategy (Papathanassis & Knolle, 2011). As a result, electronic word-of-mouth (eWOM) is currently the most significant information source influencing consumer purchasing decisions in the e-commerce industry (Zhang et al., 2016). Furthermore, as e-commerce gains popularity, the number of customer reviews per product grows rapidly (Hu et al., 2004).

Consumers have used electronic word of mouth (eWOM) extensively over the past few decades on various online platforms to acquire product feedback from prior customer experiences. However, this helpful information needs to be clarified by detecting and excluding fake reviews (Borah et al., 2020). The European Parliament defines a fake review as a positive, neutral, or negative review that is not an actual consumer's honest and impartial opinion or does not reflect a consumer's genuine experience of a product, service, or business (European Parliament, 2015).

Because consumers are greatly influenced by other shoppers' views and experiences, user-generated material, like product reviews on Amazon, can enormously shape and influence consumer purchase decisions (Nguyen et al., 2018). In 2020, the review-detection website Fakespot analyzed 720 million Amazon reviews, and the results revealed that 42 percent of those evaluations were fakes (Fakespot, 2021). Despite the commitment of Amazon to combat fraudulent reviews, nowadays, it is estimated that between 10% and 30% of reviews on websites are fake (Otero, J. 2021). Approximately $152 billion worth of goods and services were purchased globally last year as a result of fake reviews, according to the World Economic Forum analysis (McCluskey, M. 2022).

§ 1.2 Problem statement

Despite much interest from the academic world and companies, fake reviews have proven to be particularly difficult due to the various review writers, employees, fake-review marketplaces, machines, and customers (Berger et al., 2020). Besides, studies have shown that consumers also need help distinguishing fake from genuine reviews. Because of this ongoing problem, there is an urge for more transparency and information on detecting fake reviews and

their relevant features (Azimi et al., 2022). Researchers proposed various features to feed detection algorithms for a better classification system. These features are categorized into review-based features (content of review), product-based features (attributes of product), and reviewer-based features (characteristics of reviewer). As fake users can always change their identities, review-based features are most popularly used to detect deceptive online reviews (Hassan & Islam, 2021). Online reviewers have lots of time to consider whether they are being dishonest or truthful, allowing them to create more convincing textual reviews or fraudulent reviews. Due to the freedom in word choice, it might be challenging to distinguish reviews between real and fake when using linguistic and sentiment analysis features in web-based situations (Abri et al., 2020). Linguistic features consist of quantity (including an average number of words, part of speech (POS), modifiers, capital words, punctuations, and sentences in the text), complexity (including sentence/word length, redundancy, and readability), and diversity (lexical diversity, i.e., % unique words). Sentimental analysis features emotions, including positivity, negativity, and objectivity/neutrality.

Previous studies have shown that linguistic and sentiment analysis features can be used to distinguish between fake and real reviews with great accuracy (Abri et al., 2020; Bharatkumar et al., 2022). In order to distinguish between fake and real online reviews, this thesis will focus on analyzing textual features like sentiment and linguistics. Furthermore, we will investigate to what extent a machine learning algorithm can detect fake reviews using only linguistic and sentimental features in e-commerce and which linguistic and sentiment analysis features are most effective.

We will use Natural Language Processing (NLP) to extract the linguistic and sentiment analysis features. This method enables computers to process human language in the form of text and to 'understand' its whole meaning, complete with the writer's sentiment (Vaitheeswaran & Arockiam, 2016). Reading a large number of reviews takes time and effort. Therefore, it is helpful to employ techniques to summarize them automatically. Furthermore, fake reviews are posted to promote or demote the products. Hence the sentiment is usually much higher or lower when the review is deceptive (Bharatkumar et al., 2022). Therefore, opinion mining/sentiment analysis is implemented to detect a review's sentiment. Opinion mining or sentiment analysis is mining the text's behavior, opinions, and sentiments. There are two-ways sentiment analysis can be performed: lexicon-based and machine-learning approaches (Birim et al., 2022). To train a machine learning model to classify sentiment, it needs a labeled dataset based on the emotions expressed in the text. Because, for most real-life instances, there are no labeled datasets at hand to train an algorithm, this research will focus on the lexicon-based approach.

§ 1.3. Relevance

The detection of fake reviews is a well-recognized problem that has attracted significant interest from the research community. Despite the majority of online platforms having their own algorithms for detecting fake reviews (Cheng et al., 2017), these algorithms occasionally have a narrow scope and only catch 16% of posted fake reviews (Luca & Zervas, 2016). As a result, it becomes clear that new methods and algorithms must be developed. The academic world has explored different detection methods in the last couple of years. Most researchers investigated machine learning approaches to detect fake online reviews. Include supervised classification models and unsupervised classification models. Supervised methods (Saumya & Singh, 2018) learn a classifying model from labeled information (i.e., fake and real reviews). Unsupervised methods apply clustering techniques (Hu et al., 2013) and graph-based analysis (Ye & Akoglu, 2015) for fake review detection without requiring labeled data. Mostly used machine learning approaches are Support Vector Machines (SVM), Decision trees, Naive-Bayes (NB), Random Forest, Logistic regression, and Multiple Layer Perceptron (MLP) (Abri et al., 2020). Especially the Support Vector Machines and Naïve Bayes have proven to be successful in detecting fake reviews due to their ability to model complex relationships. We will further explain the machine learning approaches in Chapter 5, Model.

In order to understand the offered information, it is crucial to develop systematic methods (Nguyen et al., 2018). Consumer-created reviews of products and services are a critical driver of everyday decision-making. However, the credibility of these reviews is damaged when businesses commit review fraud, either fabricating negative reviews for the competition (vandalizing) or generating positive reviews to benefit themselves (boosting). Online reviews, as word-of-mouth communication, significantly impact how consumers choose products and services, as well as how businesses perform financially and develop new products and services. This results in unexpected incentives to generate fake reviews (Zhang et al., 2016). In one of two ways, consumers are harmed by fake ratings: either they end up paying more than they otherwise would have had the product not been deliberately overrated, or they end up purchasing a product of lower quality than the closest substitute (He, X. 2023).

Until now, Amazon has no transparency on how they combat fake reviews or how consumers can detect them based on guidelines (Hill, S. 2022). This research aims to contribute to transparency for the consumer to understand better the textual features presented in fake reviews. For both consumers and sellers, fraudulent reviews can have significant consequences.

Therefore, both sides must gain a deeper grasp of how these types of deviant communication are created.

Although research on the performance of these linguistic features in the context of textual data to detect fake reviews has been done (e.g., Abri et al., 2020; Wang et al., 2022; Lai et al., 2023; Alsubari et al., 2020), the effectiveness of sentiment-lexicons has not been explored yet. Research has mainly focused on the usage of machine learning techniques with text to detect fake reviews (Royal et al., 2023; Choi et al., 2022; Shahariar et al., 2019) or sentiment analysis to detect fake reviews (Patel et al., 2018; Aono, T. 2019; Peng & Zhong, 2014; Saumya & Singh, 2018). Reducing the quantity of linguistic and sentiment analysis features used for fake review classification is desirable. In order for the detection model to be simplified and reduce the amount of noise and uncertainty brought on by insignificant features (Abri et al., 2020). Therefore, this research will answer the following questions: How can fake reviews be detected based on linguistic and sentiment analysis features? How can sentiment analysis assist detection of fake reviews, and how do they work? Which linguistic and sentiment analysis features are most helpful in detecting fake reviews using machine learning? What are the advantages of knowing fake reviews' linguistic and sentimental features?

§ 1.4 Method

Recent advancements in Natural Language Processing (NLP), especially when combined with machine learning, have provided great results in the classification of reviews. In this thesis, we will continue building on that success. This research will not use the actual text of the review but rather extract all the linguistic and sentiment features to train the classifier to detect fake or genuine reviews. The results give new insights into the influence of specific characteristics and the impact sentiment has on the legitimacy of a review. Lexicon-based techniques will be used to research the sentiment of the reviews. Due to their success in previous scientific papers, the two lexicon-based approaches are SentiWordNet and Valence Aware Dictionary and Sentiment Reasoner (VADER) (Nguyen et al., 2018). Thanks to sentiment analysis techniques, quantitative ranking can now be obtained from textual reviews (Kauffmann et al., 2020). These approaches use dictionaries to score words based on their polarity (positive, negative, or neutral). We will use lexicons to extract the sentiment scores and research, e.g., the intensity, ambiguity, number of polarity words, modifiers, shifters, negations, and emoticons. Using NLP, we will extract the text's quantity, complexity, and diversity. NLP is a tool that enables to quantify qualitative information and to spot patterns or make comparisons (Berger

et al., 2019). The text analysis workflow for this study will look as follows: Data preprocessing, Entity (word) extraction, and measurement.

The review dataset used in the research is posted and annotated by Amazon and will be used to train the algorithm to detect fake reviews. The corpus comprises 21,000 reviews labeled as fake or real (Liev, G. 2019). To classify the reviews, we will use the Support Vector Machine (SVM) for the main test and Naïve Bayes (NB) to compare its performance. We will choose these base learners because they showed great potential to analyze fraudulent reviews in previous works (Baishya et al., 2021). Finally, we will discuss the working of the algorithms and previous success in the coming chapters.

Analyzing the supervised algorithms and lexicon sentiment analysis will bring new insights to improve the detection of fake reviews and influence of sentiment analysis and linguistic features. To the best of our knowledge, research has yet to be done regarding combining a lexicon approach with a machine learning approach for an e-commerce dataset based on only textual information. The studies of Abri et al. (2020), Vanta & Aono (2019), and Dewang & Singh (2015) all used different kinds of linguistic and sentiment analysis features to detect fake reviews. Abri et al. (2020) used only linguistic features, and Vanta & Aono (2019) used linguistic and sentiment analysis features to spot fake reviews but also used the text. Dewing & Singh (2015) used lexical features to identify the reviews like POS Tagging, Readability, and Lexical Diversity. Our main contribution will be to illustrate how textual features can be used to predict and understand fake reviews without the actual text. We will provide an overview of the most valuable and significant features to create a better understanding of detection by businesses, consumers, and researchers.

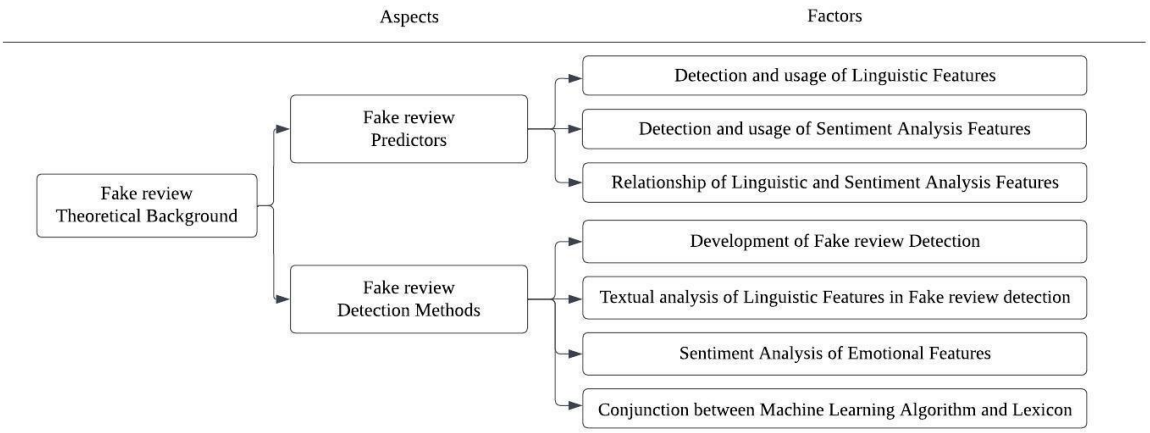
The remainder of this paper is structured as follows. After an extensive literature review in 'Theoretical Background', we present the methodology used in the present study in 'Method'. Next, we will elaborate on the data in the section 'Data.' Hereafter we will report our findings in 'Results'. Finally, the paper concludes with a discussion of the implications of our findings and general conclusions with future research ideas in the last chapter, 'Discussion'.

Chapter 2: Theoretical Background

The detection of fake reviews has grown in importance as a research topic, inspiring the creation of numerous techniques and methods. These methods include machine learning-based approaches, which utilize different features. Additionally, sentiment analysis, which involves analyzing the textual information within reviews to determine the writer's sentiment, has proven effective in detecting fake reviews (Abri et al., 2020). One approach, in particular, can potentially improve current classification techniques, lexicons (Patel et al., 2018). This technique allows researchers to calculate sentiment without the use of any pre-classified dataset and analyze review-based features (Berger et al., 2019).

This chapter aims to give a broad overview of the theoretical background, focusing on linguistic and sentiment analysis features within fake reviews and identifying fake reviews, covering the techniques and methods used for natural language processing and sentiment analysis. We will review the predictors of fake reviews in the first part of this chapter. This review gives us the necessary insight into the variables for this research and the current knowledge regarding the usage of linguistic and sentiment analysis features in fake review detection. The second part will consist of the literature on detecting fake reviews. This section aims to give the reader a clear understanding of the development and state of fake review detection. We will scrutinize textual and sentiment analysis using a machine learning vs. lexicon-based approach and the conjunction between feature extraction and detection methods. The results will be summarized and utilized in the last part of the chapter to set up the conceptual framework for the research. We will display the structure of Chapter 2 in the flowchart below:

Figure 1: Theoretical Background structure



§ 2.1. Literature Fake Review-based Predictors

§ 2.1.1. Detection and Usage of Linguistic Features

As defined in Chapter 1, Introduction, linguistic features in a text refer to the observable elements and patterns that contribute to its structure, meaning, and communication (Crossley, S., 2020). Therefore, text-related features are essential in fake review classifications (Sule et al., 2022). Furthermore, a review writer must write the reviews convincingly and consistently in order to capture readers' attention and affect a consumer's purchasing decisions. As a result, it is expected that manipulators' writing styles will differ from those of true consumers (Hu et al., 2010).

Recent developments in natural language processing (NLP) have made it possible to more accurately compute linguistic features in big datasets of reviews, which has allowed researchers to gain insight into a variety of cognitive processes, such as how people judge the quality of texts and the styles of writing (Crossley, 2020). Abri et al. (2020) applied NLP to extract different linguistic features as input for machine learning fake review detection algorithm. Their research analyzed linguistic features in the category's quantity, complexity, non-immediacy, expressiveness, diversity, informality, and specificity. In addition, they identified the number of adjectives, redundancy, lexical diversity, and pausality as the most important features to classify whether a review was fake or trustworthy.

Vanta and Aono (2019) used 24 linguistic and sentiment analysis features, including the rating. They combined the extracted features with a Bag-of-Words approach, where they vectorized all the words to calculate the importance of each word based on the number of times it occurs in the dataset. They concluded that word count, the extremity of rating, and the ratio of numerals were the most important variables in detecting fake reviews combined with using Bag-of-Words.

Dewang and Singh (2015) used Lexical features to study fake reviews. These features represent characters and different words used by the writer of a review. Using 17 different features resulted in promising results to classify reviews as fake or real. Finally, Alsubari et al. (2020) researched the identification methods of fake reviews based on linguistic features. They used features like POS count, polarity score, and authenticity score to train their model, and with success.

Several studies focused on only specific linguistic features, like Ghose and Ipeirotis (2011). They used readability as a predictor for fake review detection. Many metrics have been effective for measuring a text's readability. Based on research on readability, the Flesch Reading Ease is a helpful metric for assessing how simple it is for an individual to read a review (Ghose

& Ipeirotis, 2011). A formula produces a score for readability. It was created based on a mathematical model that evaluated how easy it was for subjects to read various text samples. Also, Hu et al., 2011, found that readability is a significant predictor for manipulations of online reviews.

Concluding, Natural Language Processing opened the doors to research linguistic features by extracting them from the text. Previous research has exposed several linguistic features promising to detect fake reviews, but there is always room for improvement.

§ 2.1.2. Detection and Usage of Sentiment Analysis Feature

The sentiment of word-of-mouth (WOM) is critical for consumers to share their experiences and evaluate products, as WOM communication can be positive or negative. Positive WOM (PWOM) is believed to stem from satisfactory experiences. In contrast, negative WOM (NWOM) is often driven by motives and needs associated with a negativity bias, as negative information is easier for consumers to perceive (Vázquez-Casielles et al., 2013). These emotions are presented as sentiment analysis features in text analysis (Bandhakavi et al., 2016).

Business owners and (potential) consumers post fake reviews for different purposes. Business owners often create fake consumer identities to post positive reviews to promote their products/services or negative reviews to demote competitors' products/services (Luca & Zervas., 2016). Meanwhile, individual (potential) consumers post positive and negative fake reviews for economic or personal reasons (Sigala et al., 2017). Deviant consumers may create fake identities and post negative fake reviews when they are unsatisfied with a company's products/services or positive fake reviews to support the business of their friends or family or receive gifts (Hunt, 2015). Regardless of the source, fake reviews can reduce the credibility and value of online reviews (Hunt, 2015). According to the theory of negativity effects, negative information is more straightforward for consumers to perceive than positive information; therefore, negative information can substantially influence purchase decisions. For this reason, stakeholders need to know the sentiment of reviews (Gavilan et al., 2018).

Knowing the sentiment in a review gives us insight into a writer's intention. However, it has also been identified as a reliable predictor of whether a review is fake or not (Wang et al., 2020). General emotional features that can be extracted from a text are contextual features, sentiment features, polarity shifters, modifiers, and negations. These features help dissect the text in numerical information that is perfect for feeding a machine learning algorithm and has had promising results in detecting fake reviews (Bandhakavi et al., 2023). Besides, there are previous papers investigating the relationship between fake reviews and sentiment analysis

features; the way these features are harvested and combined with other features needs to be more enlightened.

§ 2.1.3. Relationship of Linguistic and Sentiment Analysis Features

Jung et al. (2020) established a significant impact of quantitative and non-quantitative characteristics of customer reviews on a consumer's decision-making process. Because of evidence of significant impact, they confirmed that product-related quantitative and textual information affects e-commerce and emphasized the significance of taking consumer text reviews seriously. (Choi et al., 2022). Various features, like linguistic features, readability, and redundancy, have been proposed for detecting reviews spam. Azimi et al. (2022) also concluded that redundancy and readability are important features in detecting fake reviews. In addition, researchers have recently included sentiment analysis features to strengthen the detection methods' accuracy (Alsubari et al., 2020).

Ghose et al. (2011) used text analysis to identify the crucial aspects of the review. To find linguistic features with strong predictive power in determining a review's usefulness and economic impact, they conducted an analysis at the lexical, grammatical, semantic, and stylistic levels. Subsequently, Hu et al. (2012) were the first researchers to believe that using certain linguistic features in combination with sentiment would be an important step in figuring out the effects of fake reviews.

Besides, Jindal and Liu (2008) noted that fake reviews are repetitive most of the time because their authors had neither bought nor used the products they were reviewing. Therefore, to explain some of the product's features or point out some of its disadvantages, they needed help. So, in order to determine whether a review is fake, it is crucial to compare how the reviews are written (Wang et al., 2022). This assessment will be easier to conclude by applying a feature set of linguistic and sentiment analysis features.

§ 2.2. Literature Fake Review Detection Methods

§ 2.2.1. Development of Fake Reviews Detection

Detecting fake reviews is now a critical field of research for academics and practitioners, as it is imperative to strengthen the reliability and validity of website posts (Alsubari et al., 2021). In this context, developing effective algorithms using machine learning techniques has gained prominence to enable the automated detection of fake reviews. Algorithm detection involves building a machine-learning classification model and using similarities to detect fake online reviews (Shukla et al., 2019). To detect fake reviews, machine learning approaches, such

as supervised and unsupervised, have been extensively used (Crawford et al., 2017). In addition, various machine learning classifiers, such as Support Vector Machines (SVM) or Support Vector Classifier (SVC), Decision trees, Naive-Bayes (NB), Random Forest, Logistic regression, and Multiple Layer Perceptron (MLP), have been used (Abri et al., 2020).

The machine learning algorithms each offer different approaches. Decision Tree (DTC) constructs a tree-like model using feature thresholds, providing an easily interpretable solution. Random Forest (RF) combines multiple decision trees to improve accuracy and prevent overfitting. Support Vector Machine (SVM) and Support Vector Classifier (SVC) have been proven helpful in binary classification, seeking to separate different classes. Naive Bayes (NB) is a probabilistic classifier assuming feature independence. Logistic Regression (LR) is a way to determine the likelihood of a feature belonging to a particular group using a math formula called a logistic function (Elmogy et al., 2021). Finally, the Multiple Layer Perceptron (MLP) is a multi-layer neural network that excels at capturing complex relationships, using multiple layers to find important features and identify patterns in data (Abri et al., 2020). In Table 1, the result of numerous fake review detection papers are listed:

Table 1: Prior Research on Fake Review Detection Methods

Authors	Dataset	Used Features	Machine Learning Algorithm	Accuracy Score
Salminen et al. (2022)	OpenAI written Amazon Dataset	Review-based and Product-based features	NB SVM	95.82% 95.82%
Erlmurngi & Gherbi (2018)	Movie reviews Dataset	Sentiment, Text, and Reviewer-based features	NB SVM	70.9% 76%
Baishya et al. (2021)	Amazon Dataset	Rating, Text, Time, and Reviewer ID	NB SVM RF	89% 94% 91%
Elmogy et al. (2021)	Yelp Dataset	Product-based features and Caps-count, Punct-count, Emojis, and Text	NB SVM LG RF	85.82% 86.9% 86.89% 86.85%
Moqueem (2023)	Amazon Dataset	Text, Rating, Verified, and Date	NB SVM RF	79.2% 78.6% 75.8%
Birim et al. (2022)	Amazon Dataset	Review-based (e.g., Length, sentiment), reviewer-based features (e.g., Verified purchase, Rating), and Text	DT RF SVM	78.17% 80.4% 78.76%
Pendyala, A. (2019)	Consumer review Dataset	Reviewer ID, Rating, Verified purchase, Sentiment, Text, Part-of-Speech, Emoticons	NB SVM, RF	84% 81% 79%
Shahariar et al. (2022)	Yelp Dataset	Text	NB SVM	91.75% 91.73%
Narayana Royal et al. (2023)	Amazon Dataset	Text, Part-of-Speech, and Word-count	NB SVM	69% 80%

Table 1 shows that the Support Vector Machine (SVM) and Naïve Bayes (NB) are among the most common methods. They show excellent performance and are useful in detecting fake reviews. The accuracy (i.e., correct predictions) variates from 69% to 96%. SVM

is great at drawing a clear line between genuine and fake reviews. SVMs work by finding the best possible separation between the two classes, making them highly effective for detecting fake and real reviews. On the other hand, NB is an effective algorithm, especially for tasks involving text, like detecting fake reviews. It assumes that the features are independent (Hossain, F. 2019).

Narayana Royal et al. (2023) used NB and SVM to detect fake reviews on an Amazon dataset. They chose these machine learning algorithms due to their success in previous studies in identifying fake reviews due to their classification technique. Their research achieved an accuracy of 80.1% with SVM and 68.7% using NB.

In summary, various techniques have been employed to detect fake reviews, each with its strengths and limitations. In most cases, SVM was the most successful machine learning algorithm, next to NB. For this reason, our research will implement these methods to detect fake reviews. The following two sections will clarify the method to extract the features to feed the algorithms.

§ 2.2.2. Textual Analysis of Linguistic Features in Fake Reviews Detection

Prior studies have employed various methods, including linguistic analysis, to detect fake online reviews. Some studies have focused on identifying linguistic cues, such as affective, cognitive, social, and perceptual, to differentiate fake reviews from genuine ones. Prior research has also identified general linguistic cues, such as specific words, structural properties, functional words, and punctuations and tenses, as potential indicators of fake online reviews (Abri et al., 2017). In Table 2, various linguistic analyses that use different features to detect fake reviews are presented:

Table 2: Prior Research on usage of Linguistic Features for Fake Review Detection

Authors	Dataset	Used Features	Feature selection	Classifier	Accuracy Score
Abri et al. (2020)	Restaurant Dataset	Quantity, Complexity, Non-Immediacy, Expressiveness, Diversity, Informality, and Specificity	RF and Recursive-Feature Elimination	SVM NB RF LR DTC MLP	77.27% 73.63% 75.45% 73.67% 73.63% 79.09%
Ghose & Ipeirotis (2011)	Amazon Dataset	Product-based, Review-based, Reviewer-based, Readability, and History-based features	RF	SVM	84.68%
Hu et al. (2012)	Amazon Dataset	Price, Rating, Readability, Sentiment, Words, Length, and helpful votes	Linear Regression		
Choi et al. (2022)	Naver shopping Dataset	Sentiment and Quantitative features, Source, Text, and Rating	BERT	SVM RF LG NB DT	85.1% 83.8% 83.6% 82.1% 81.1%

Alsubari et al. (2020)	Yelp Dataset	Authenticity, Sentiment, Subjective, Part-of-Speech, Rating, and Word count	LIWC	RF DT	94, 96% 94, 96%
Wang et al. (2022)	E-commerce Dataset	Length, Relativity, Word repetition, Sentiment, Word count	Recursive-Feature Elimination	SVM NB	72, 79% 72, 79%
Dewang & Singh (2015)	Hotel Dataset	Writing styles, Length, Word count, Lexical Density, Part-of-Speech, Syntactic features	WEKA	NB DT	81.24% 84.63%

As shown in Table 2, most linguistic features can be assigned to the category's quantity, complexity, and diversity. When analyzing transcribed speech, linguistic features are often the focus. This paper implements linguistic features to investigate their effectiveness in enabling the detection of fake online reviews.

Previous research on identifying fake reviews has demonstrated that relying solely on linguistic features did not produce satisfactory results, with accuracy varying from 55% to 68%. Performance was greatly improved by around 20% when extra behavioral features like sentiment were incorporated. The performance and accuracy of detection would improve with the extraction of more features (Wang et al., 2020).

In short, linguistic features are very effective and useful bases for classifying fake and truthful reviews. Moreover, in combination with machine learning approaches like Support Vector Machine and Naïve Bayes, they can give insight into the detection of features of fake reviews that are deceitful for the consumer. Overall, studies demonstrate the potential of linguistic features in detecting fake online reviews.

To improve the performance and accuracy of detection significantly, further research needs to be executed to examine the impact of incorporating additional features. All of the mentioned research also used the text as input for the classifier to detect fake reviews, besides Abri et al. (2020). However, Abri et al. (2020) did not include sentiment analysis features as potential variables. This research will fill the gap of using linguistic features, without the text as input, to detect fake reviews, and will include the sentiment of a review which we will cover in the section. Based on these previous findings, we will use the linguistic features corresponding to the category's 'quantity', 'complexity', and 'diversity'.

§ 2.2.3. Sentiment Analysis of Emotional Features

A major research area has emerged around extracting the best and most accurate method while also classifying the customer's posted review text as negative or positive. The two main techniques for determining if a review's sentiment is positive or negative are machine learning and lexicon-based (dictionary) approaches (Bandhakavi et al., 2018). A machine learning method trains a classifier on a labeled dataset to predict its sentiment (Xhymshiti, M., 2020). A

lexicon-based method examines the sentiment of a review using a dictionary of terms and assigns a polarity label to each word (e.g., positive, negative, or neutral). The dictionary-based analysis approach is a technique that quantifies reviews by matching preprocessed review data to an emotional dictionary that has been pre-build. To analyze qualitative data, such as user review text, researchers commonly use dictionary-based analytic approaches (Choi et al., 2023). In addition, manipulators are likely to apply specific persuasion techniques in an attempt to generate reviews that consumers will believe and act upon (e.g., polarity shifters or intensity modifiers). In Table 3, the most prominent sentiment analysis research is presented using techniques like lexicons and machine learning:

Table 3: Prior Research on usage of Sentiment Analysis for Fake Review Detection

Authors	Dataset	Fake reviews	Used Features	Sentiment-Analysis Approach	Classifier Fake Review	Accuracy Sentiment	Accuracy fake reviews
Nguyen et al. (2018)	Amazon Dataset	NO	Text	Lexicon VADER, SentiWordNet, Pattern		83% 80% 69%	
Mitra, A. (2020)	Movie reviews Dataset	NO	Text	Lexicon VADER SentiWordNet (NLTK)		80%	
Peng & Zhong (2014)	Resellerrating Dataset	YES	Text, rating, vocabulary, number of words/sentences	Lexicon MPQA SentiWordNet	DSP		59.6% 61.4%
Wang et al. (2020)	Yelp Dataset	YES	Lexical, Sentiment and tekst-Based features	Lexicon SenticNet	RF LR DT NB SVM		76.96% 76.76% 68.33% 69.87% 78.88%
Saumya & Singh (2018)	Amazon Dataset	YES	Rating, Review sentiment, Comments, and Helpful votes	Machine learning	RF SVM		91% 65%
Taqiuddin et al. (2021)	Steam Dataset	YES	Sentiment and Reviewer-based features	Lexicon TextBlob	SVM		81%
Xhymshiti, M. (2020)	Bol.com Dataset	NO	Text	Lexicon Pattern SVM		65% 68%	
Machova et al. (2022)	Reddit Dataset	YES	Sentiment and Reviewer-based features	Lexicon Afinn	SVM		84%
Gupta & Mandal (2017)	News Dataset	NO	Text	Lexicon-based SVM NB		81% 89% 88%	
Srivastava et al. (2022)	Tripadvisor Dataset	NO	Text	Lexicon Afinn VADER		86% 88.7%	
Patel et al. (2018)	Amazon Dataset	YES	Text, Polarity score	Lexicon VADER			
Vanta & Aono (2019)	Yelp Dataset	YES	Quantative features, Rating, Part-of-Speech, Text, Ratios	Lexicon Liu & Hu	SVM, MLP		77.3% 77.5%

As shown in Table 3, some of the most popular lexicons are VADER and SentiWordNet. For example, Srivastava et al. (2022) concluded that the VADER-lexicon is the most accurate for determining the sentiment of a review in a Tripadvisor dataset, with an accuracy of 88.7%. Also, when using sentiment lexicons combined with a machine learning classifier, the results vary between 60% and 84% accuracy.

Peng & Zhong (2014) used the MPQA-Lexicon and SentiWordNet lexicon to detect spam reviews. They calculated the Sentiment score and incorporated this variable as input for the detection algorithm. They found that the sentiment of a review is an important predictor of recognizing a fake or real review. In comparison to the usage of rating or word count alone, the accuracy of the classification models presented in this paper increased from an average of 75% to 85%.

Concluding, there has been some research into the working of lexicons regarding fake review detection. Researchers have proven that the features are effective and represent an important predictor, but more in-depth research is needed.

§ 2.2.4. Differences between machine learning and lexicon

Different methods have been used by researchers to address sentiment analysis problems. One method is unsupervised machine learning, which uses classification algorithms to categorize data based on related sentiment. A set of training data must be used for supervised learning to learn. Naive Bayes (NB) and Support Vector Machine (SVM) are popular supervised machine learning techniques for classifying texts based on sentiment. There is also more scientific interest in semi-supervised learning models like Lexicons. The Lexicon dictionaries allow researchers to compute sentiment scores and label data without pre-labeling.

Gupta and Mandal (2017) compared the working of supervised machine learning and lexicon-based classification. They concluded that supervised machine learning approaches SVM and NB are more successful in classifying a sentence as positive, negative, or neutral than a lexicon-based approach (93% accuracy vs. 81% accuracy).

Although supervised learning methods perform at a higher accuracy rate than lexicon-based approaches, the main problems of supervised and unsupervised learning are well-balanced by semi-supervised learning. Until recently, (un)supervised machine learning techniques have only used the basic variables, such as part of speech, to classify the sentiment. Ignoring factors like the interactions between several features reduces the classification effect (Wang et al., 2020). When using a lexicon-based approach, the dictionary will take these interactions into account and allows researchers to extract certain related features from a text, like polarity shifters (changes direction of the emotion, like 'not'), modifiers (intensify the text, like 'very'), negations, and emoticons. For this research, we will use a lexicon-based approach because they can calculate sentiment without using a pre-labeled dataset and can extract manipulators like polarity shifter from the text.

§ 2.2.5. Conjunction between Machine Learning Algorithm and Lexicon

Prior studies used two approaches to identify fake reviews: algorithm detection and linguistic/sentiment analysis. Both approaches have shown great results in identifying fake reviews. These approaches have been proven to be even more successful in identifying fake reviews in combination with each other (Shukla et al., 2019).

Although machine learning techniques are more reliable than traditional statistical or manual detection, existing methods have some areas that need improvement. For example, a significant problem in machine learning is finding definitive clues to classify reviews as real or fake. To find definitive clues, a common approach is to use feature extraction to identify a more accurate feature set that will better classify reviews (Crawford et al., 2015). With this, it is necessary to design automated fake-review detection approaches using sentiment analysis and natural language processing (Anusha & Prasad, 2020). Berger et al. (2019) concluded that the best approach is to train a supervised machine learning tool by using natural language processing to understand the linguistic relationship in the sentence, as this will result in better insights.

A different study using the Support Vector Machine technique and lexicon-based features for review detection and sentiment analysis concluded that combining the two increased accuracy from 60% to 84.6%. Therefore, combining lexicon features with a fake review detection model can make up for each feature's weaknesses to produce more accurate results. (Taqiuddin et al., 2021). Also, Vaitheeswaran & Arockiam, 2016, concluded that using a machine learning classifier in conjunction with emotional features and contextual text analysis will produce more accurate results than prior research.

So far, a few studies have compared Machine learning algorithms and Lexicon-based techniques. E.g., Nguyen et al. (2018) looked at the difference between the two sentiment classification methods, using SVM, GB, and LR algorithms and VADER, Pattern, and SentiWordNet lexicons on an Amazon dataset.

Until now, there is scientific evidence that sentiment analysis and machine learning techniques can help identify fake reviews. However, no research was executed to see the combination of these two using lexicon-based sentiment classifiers in combination with linguistical features. As Berger et al. (2019) explained, defining the best feature set by using natural language processing to feed a supervised machine learning tool gives better insights and will result in potential improvements.

§ 2.3. Conceptual Framework

Most research used e-commerce datasets, as shown in Tables 1, 2, and 3. Because fake reviews significantly impact the success of e-commerce retailers, their influence on the decision-making process of consumers, and previous successes, an Amazon dataset will fit perfectly regarding this research. Various scientific papers used the dataset, e.g., Bharatkumar et al., 2022; Alsubari et al., 2023; Diav et al., 2020; Narayana Royal et al., 2023. The dataset will receive further elaboration in Chapter 3, Data. After selecting the dataset, we will extract the linguistic and sentiment analysis features. As shown in research into detecting fake reviews using linguistic features (Abri et al., 2020; Choi et al., 2022; Alsubari et al., 2020), we will use natural language processing to extract the input for the detection method. The code will extract the features related to the quantity of the text per review, including the number of words, number of sentences, capital words, punctuations, and part-of-speech (number of nouns, verbs, adjectives, and adverbs);

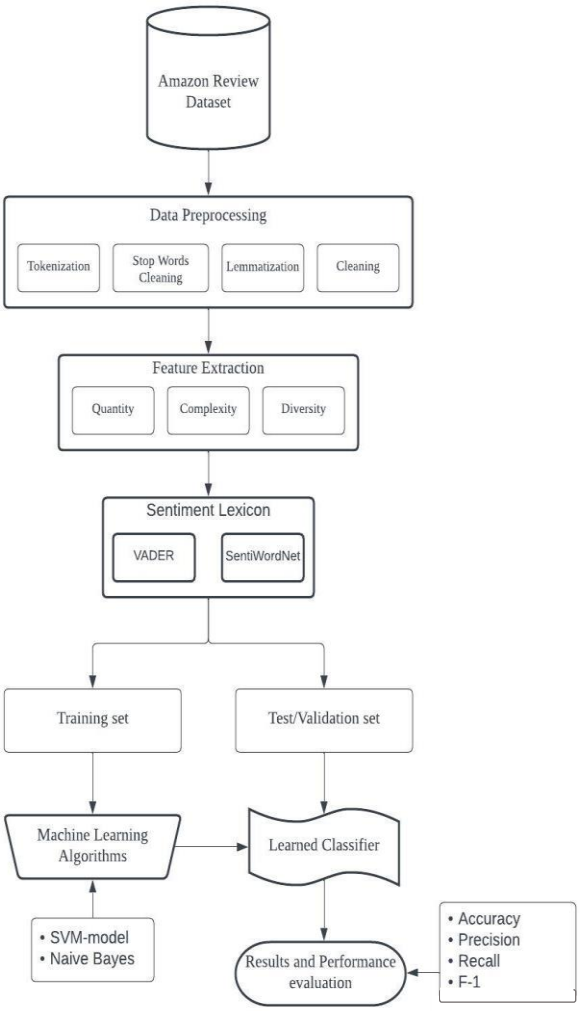


Figure 2: Conceptual Framework

the complexity of the text per review, including average word length, average sentence length redundancy score and readability score; the diversity of the text per review, including lexical diversity.

Abri et al. (2020) already proved that redundancy and lexical diversity were of significant value in detecting fake reviews. Redundancy refers to the unnecessary repetition or inclusion of information, words, or phrases that do not contribute to the overall meaning or clarity of the text (Crossley, S. 2020). Lexical diversity, on the other hand, measures how many different words are presented within a review (Dewang & Sing, 2015).

We will use the lexicon-based approach to extract the sentiment analysis features because lexicons do not need labeled datasets and can extract more data than machine learning algorithms (Peng & Zhong, 2014). Because of previous successes and its prominence in prior research, as presented in Table 3, this research will incorporate the VADER and SentiWordNet

lexicons to extract the sentiment analysis features per review, including sentiment score, positivity, negativity, and objectivity, number of words in different sentiment categories (positive, negative, objective/neutral), sentiment score per word.

Finally, the most significant features to detect fake reviews will be selected from the extracted features. We will use Recursive Feature Elimination (RFE) to select the most promising feature set because of its previous success by Wang et al. (2022) and Abri et al. (2020). RFE trains the selector based on the Random Forrest algorithm. Only statistically important features are retained as they contribute more to model performance (D'Agostino, A. 2021). Hereafter, we divide the dataset into training and testing/validation sets (standard 80-20% distribution). Like concluded in multiple papers regarding fake review detection, SVM and NB are among the most promising classifiers, as presented in Table 1. Because of this, we will utilize the supervised learning algorithms SVM and NB to train the model on the training data and assess its performance on the testing data. Further elaboration on the working of the algorithms will be in Chapter 3, Method.

Chapter 3: Method

This chapter will discuss the method utilized to classify and detect fake reviews. First, we will look at the Bag-of-Words method to preprocess the text. Hereafter we elaborate the natural language processing method. Next, we will explain the working of the lexicons and the working of the classifiers, and we will finalize the chapter with the metrics to evaluate the performance of the detection method.

§ 3.1. Natural language processing

We will use natural language processing (NLP) to extract the linguistic features. According to Berger et al. (2019), the best prediction performance to classify is obtained when a large number of textual features are combined with machine learning. Therefore, it is more about how a set of features could be used to predict a result rather than any individual feature. We will mine the features at an entity (word) level. The advantage of this approach is that we can extract a large number of entities and can combine them with dictionaries to extract linguistic styles and sentiments later on. It enables us to investigate both what was written (the words' content) and how it was written (the writer's style) (Berger et al., 2019). We will be using a Bag-of-Words (BoW) approach to create sequences of entities per review which will be considered to extract the linguistic and sentiment analysis features. The BoW method is a simplified text representation technique that ignores grammar and word order, focusing only on the frequency of individual words in a document (Hamouda et al., 2011). In fake review detection, the presence or absence of specific words can be indicative of fake or legitimate messages, therefore we will utilize this approach to count the Part-of-Speech words and modifiers (Hajek et al., 2020). Later, we will also implement this approach to extract the sentiment of the reviews as this approach works perfectly with lexicons, but we will discuss this in paragraph 3.2.

We will use entity extraction to create a large collection of entities (words) for training the prediction model and as input for dictionaries to extract more complex kinds of textual expressions, such as writing styles and sentiment. Linguistic-type entities are a kind of entity extraction, such as part-of-speech tagging, which gives each entity a linguistic tag (such as a noun, verb, or adjective) (Abri et al., 2020). Additionally, we will use count measures. The frequency of each entity's occurrence and the relationships between entities have been measured by other researchers using count measures (Borah & Tellis, 2016).

To extract the linguistic features, we will use Linguistic-based NLP tools in Python libraries (Berger et al., 2019). We will extract linguistic features consisting of quantity

(including an average number of words, part of speech (POS), modifiers, capital words, punctuations, and sentences in the text), complexity (including sentence/word length, redundancy, and readability), and diversity (lexical diversity, i.e., % unique words). In Chapter 4, Data, we will extensively explain the used libraries, including the concrete features we extract.

§ 3.2. Working of the Lexicons

As concluded from Chapter 2, Theoretical Background, sentiment dictionaries such as SentiWordNet (Peng & Zhong, 2014) and VADER (Nguyen et al., 2018) will be utilized to extract the sentiment of the text. Azimi et al. (2022) study already suggests that emotion is an important predictor of identifying fake reviews. As introduced in paragraph 3.1, we will be using the Bag-of-Words approach to utilize the lexicons. The big advantage of BoW is that it enables us to capture sentiment information and identify which words are contributing to the sentiment classification (Augustyniak et al., 2014). The use of BoW in combination with lexicons is based on the hypothesis that individual words can be considered as a unit of opinion information and therefore may be clues to review sentiment and subjectivity (El-Din, 2016). Bag-of-Words-lexicon methods for sentiment analysis feature extraction are much faster than a supervised approach to sentiment classification while yielding similar accuracy (Hamouda et al., 2011). We will use the BoW-lexicon model to calculate the sentiment scores and count the number of polarity shifters, negations, negative words, objective/neutral words, and positive words.

The disadvantage of the lexicon approach for sentiment analysis is that it uses a "bag of words" approach. This implies that the order of the words does not matter but only relies on the cooccurrence of a word of interest. One limitation of the bag-of-words model is that it does not capture the context of the words. It treats each word as an independent feature and does not consider the relationships between words or the grammar of the language (Augustyniak et al., 2014; Almestekawy & Abdulsalam, 2019). To further elaborate on our model, we will explain the working of these lexicons:

§ 3.2.1. VADER-Lexicon

We will use the VADER lexicon for sentiment analysis in linguistic models. Positive and negative terms, as well as the degree of polarity, have explicitly been matched in the lexicon for sentiment analysis. With more than 9000 lexical characteristics (words), VADER was created exclusively to measure sentiment. Four datasets have been used to build and validate VADER: Social Media Text, Movie Reviews, Amazon Product Reviews, and NY Times

Editorial (Hutte & Gilbert, 2014). VADER produces output from four sentiment metrics: positive, negative, neutral, and compound. The compound (text's polarity) score ranges between -1 and 1. Polarity indicates either positive when it is above zero or negative polarity when it is less than zero. Because VADER is open-source and part of the NLTK package in Python, we will be able to directly use it on unlabeled text data. Using this method, we can detect emotional polarity and intensity. It is a sentiment analysis model that can evaluate a text by taking its emotions, positive/negative polarity, and intensity into account (Srivastava1 et al., 2022).

§ 3.2.2. SentiWordNet-Lexicon

One of the largest English lexical dictionaries for sentiment analysis and mining opinions is called SentiWordNet (Nguyen et al., 2018). It is one of the most basic and widely used lexicons for sentiment analysis. There are around 3300 words in it, each with its own polarity score. The freely available lexicon operates on the database provided by WordNet. WordNet is a lexical database comprising English words grouped as synonyms into what is known as synsets (Peng & Zhong, 2014). SentiWordNet's "synsets" are groups of terms with similar meanings that are assigned a score between 0 and 1. In addition, every synset is associated with a positivity score, negativity score, or objectivity (neutrality) score. The classifier committee's consensus on whether to assign a term a positive or negative label is reflected in the scores. (Patel et al., 2018). We will utilize this method because one of our main goals is to form a new set of features to detect fake reviews. Lexicon-based approaches fit perfectly for this task because they can classify words on an entity level. We will extract the following sentimental analysis features emotions, including positivity, negativity, and objectivity/neutrality.

§ 3.3. Working of the Classifiers

Because of the previous successes described in Chapter 2, Theoretical Background, we will use the ML algorithm Support Vector Machine (SVM) and Naïve Bayes (NB) to build a detection model for fake reviews.

§ 3.3.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that aims to find an optimal hyperplane (separator) to separate different classes within a labeled training dataset. In the context of fake review detection, SVM can effectively classify reviews as either fake or real based on extracted features (Abri et al., 2020). The key idea behind SVM is to transform the input data into a higher-dimensional feature space, where a linear separation can be achieved.

SVM accomplishes this by using a kernel function, which maps the data points from the original input space to a higher-dimensional space (Poonguzhali et al., 2022). In simpler terms, SVM is a way to rearrange data points so that they can be separated by a straight line (2-dimensional space) or a curved surface (3-dimensional space). The function can make the data points spread out in different directions, making it easier to tell them apart (Salminen et al., 2022).

The process of training an SVM model involves finding the optimal hyperplane (best separator) that maximizes the margin, i.e., the distance between the hyperplane and the nearest data points of each class. The data points that lie closest to the hyperplane are called support vectors, as they play a crucial role in defining the decision boundary (Taqiuddin & Bachtiar, 2021). To classify new, unlabeled instances, SVM maps them into the same feature space as the training data. The decision boundary, represented by the hyperplane, is used to assign class labels based on which side of the hyperplane the instances fall (Poonguzhali et al., 2022). In our case, if an instance falls on one side of the hyperplane, it is classified as a fake review, while if it falls on the other side, it is classified as a real review.

SVM has two key benefits that make it suitable for fake review detection. Firstly, it offers faster processing compared to many other machine learning algorithms. Secondly, SVM can achieve good performance even when the training sample size is relatively small (a couple of thousand), making it well-suited for our scenario where access to labeled data is limited (Narayana Royal et al., 2023). In this thesis, we will utilize the linear Support Vector Classifier (SVC), which is a variant of the SVM method. SVC is specifically built to classify groups and uses the same method as described before. By employing the SVC in Python, we will train a classification model that determines the "best fit" hyperplane for our dataset, enabling accurate categorization of reviews as either fake or real (Poonguzhali et al., 2022).

§ 3.3.2. Naïve Bayes (NB)

Naïve Bayes is a machine learning algorithm that operates based on Bayes' theorem, which allows us to "invert" conditional probabilities. This theorem enables us to calculate the probability of an event given that another event has already occurred (Erlmurngi & Gherbi, 2018). In the context of fake review detection, NB is employed as a generative learning algorithm that models the distribution of inputs for a specific category, namely fake or real reviews. The underlying assumption of Naïve Bayes is that the features of the input data are conditionally independent given the category or presence of any other feature. This assumption allows the algorithm to make fast and accurate predictions (Birim et al., 2022). By assuming conditional independence, Naïve Bayes simplifies the computation of probabilities by treating

each feature as if it contributes to the final prediction independently. Naïve Bayes utilizes Bayes' theorem, which states:

$$P(y|X) = (P(X|y) * P(y)) / P(X)$$

Where:

P(y|X) is the probability of class y given the input X

P(X|y) is the probability of input X given class y

P(y) is the probability of class y occurring

P(X) is the probability of input X occurring

Naïve Bayes assumes that the input features X are conditionally independent given the class y. This assumption allows us to simplify the calculation of P(X|y) by breaking it down into the product of individual feature probabilities:

$$P(X|y) = P(x_1|y) * P(x_2|y) * ... * P(x_n|y)$$

Where:

$x_1, x_2, ..., x_n$ are the individual features of X

To classify a new input X, Naïve Bayes calculates the probability of each class given X and selects the class with the highest probability (Alsubari et al., 2021; Webb et al., 2010).

The choice of Naïve Bayes for this research is driven by its simplicity and suitability for fake review detection (Salminen et al., 2022). In this study, we will employ the Gaussian Naïve Bayes algorithm, which is a variant of the Naïve Bayes classifier specifically designed for continuous variables following Gaussian (normal) distributions (Abri et al., 2020), which fits perfectly for the presented features in our dataset.

The Gaussian Naïve Bayes model is fitted by estimating the mean and standard deviation for each class. These parameters capture the central tendency and variability of the features within each class (Elmogy et al., 2021). By using these statistics, the model can calculate the likelihood of a review belonging to the fake or real review class based on its feature values.

3.3.3. Application of the Machine Learning Classifiers

In order to train the classification model, we will utilize a set of labeled data. The reviews in the Amazon dataset are pre-labeled as either fake (i.e., __label1__ in the dataset) or real (i.e., __label2__ in the dataset) (Garcia, L. 2018). The SVC and NB algorithm will train to learn how to recognize these labels based on patterns within the data. Once the classifier is trained using the training review dataset, we will test it using a separate test dataset.

To evaluate the effectiveness of the classification model, the training set will consist of 80% (16,800 reviews) of the total dataset, while the remaining 20% (4,200 reviews) will be used for testing purposes.

The extracted linguistic and sentiment analysis features, as presented in the following chapter Data, will be utilized using the two machine learning techniques. Using the 'sklearn' package, the SVM model (from sklearn.svm import SVC) and NB classifier (from sklearn.naive_bayes import GaussianNB) will be loaded into the Python-file. Support Vector Classifier, abbreviated SVC, is perfectly capable of performing multi-class classification on a dataset. Gaussian NB algorithm learns based on features in decimal form. Gaussian NB needs continuous features and works perfectly for the classification of text. In the next paragraph, 4.4 Measures, we will discuss the measures to evaluate the performance of the fake review detection algorithms. We will elaborate on the performance and results of these machine learning algorithms in Chapter 5, Results.

§ 3.4. Measures

When the most promising features are extracted and selected, and we train the classifier, we will calculate its performance by several statistics. These statistics are regularly used in fake review research (Abri et al., 2020; Salminen et al., 2022; Moqueem et al., 2023). The metrics we will use to evaluate the performance of the lexicon-ML-approach on the Amazon fake reviews dataset are:

- *Confusion matrix*: This table displays the amount of accurate predictions made by the models, true positive, true negative, false positive, and false negative. It will be used to calculate several metrics, including accuracy, precision, and recall. The findings provide characteristics as follows: (Fattahi et al., 2015)
    1) True positive (TP): represents the positive reviews that were correctly predicted by the model as positive in the testing data.
    2) True negative (TN): represents the negative reviews that were correctly predicted by the model as negative in the testing data.
    3) False positive (FP): represents the negative reviews that were incorrectly predicted by the model as positive in the testing data.
    4) False negative (FN): represents the positive reviews that were incorrectly predicted by the model as negative in the testing data.

    (Abri et al., 2020)

- *Accuracy*: This indicator counts how many of the model's predictions were accurate. It is derived by dividing the overall predictions by the predictions that were accurate.
    - *ACCURACY = (TN+TP/(TN+FP+FN+TP)*
- *Precision and Recall*: The proportion of true positives among all positive predictions is the measure of precision, while recall assesses the percentage of accurate positive predicts among all positive predictions.
    - *PRECISION = TP/(TP+FP)* &
    - *RECALL = TP/(TP+FN)*
- *F1 score*: It provides a single statistic by showing the weighted average of precision and recall, balancing the two values. It is determined by:
    - 2 * (*PRECISION * RECALL*) / (*PRECISION* + *RECALL*)

(Vanta & Aono, 2019). We will use these metrics in Chapter 6, Results, where they will show the performance and help to evaluate our model.

Chapter 4: Data

In this chapter, we will outline the analysis of the data. First, we describe the raw data collection, exploration, and preprocessing of the data set in paragraph 4.1. Then, we present our data cleaning steps and variable operationalization steps. In paragraph 4.2, we explain how we used the cleaned data to develop the linguistic and lexicon-based sentiment analysis methods to extract the features. This section will explain the extract of all features to later feed to the classifier. Next, we present the final dataset in paragraph 4.3 and assess the significance of the features by comparing the fake and non-fake reviews using t-tests. This chapter will display the set of features that will be used in the research to detect fake reviews. To finalize our research, paragraph 4.4 will discuss the last step, feature selection. We will explain this final step to act as a bridge between the data and the following chapter, Results.

§ 4.1 Data Collection, Exploration, and Preprocessing

In order to utilize a supervised machine learning approach based on a linguistic and sentiment analysis feature set, we collect a labeled dataset consisting of reviews classified as fake or real. The efficiency of fake review detection methodologies depends upon the labeled data used for training, the appropriate selection of features, and the data mining techniques employed for detection (Patel & Patel, 2018).

This study utilizes a pre-collected and labeled dataset from Amazon, comprising 21,000 reviews (10,500 fake and 10,500 real). The creators of this dataset label the reviews based on the wording within the text (Narayana Royal et al., 2023). We use the dataset "amazon_reviews", which is openly available on Kaggle (Url: https://www.kaggle.com/datasets/lievgarcia/amazon-reviews). The dataset was used in various scientific papers, e.g., Bharatkumar et al., 2022; Alsubari et al., 2023; Diav et al., 2020; Narayana Royal et al., 2023. The dataset consists of the following variables:

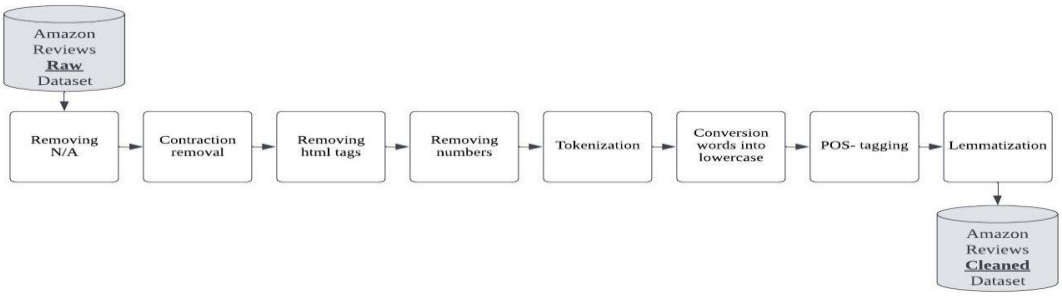
Variables	**Description**
Label	The label whether the review is real or fake
Rating	The Star-rating of the review
Verified Purchase	The label whether the reviewer bought the product or not
Product Category	The Category of the product as presented on Amazon
Product ID	The unique product ID of the review
Product Title	The name of the product
Review Title	The title of the review written by the reviewer
Review Text	The text of the review written by the reviewer

As shown in Table 4, the dataset incorporates various variables, including the star rating, verification status, and review text. With a maximum rating of five stars, the reviews had an average rating of 4.13. Additionally, verified purchases made up 55.7% of the total dataset. Moreover, the reviews are evenly spread throughout 30 product categories (such as kitchen, PC, and health), each containing 700 reviews.

§ 4.1.2 Data Preparation and Variable Operationalization

Text data, which is often unstructured, requires preprocessing to make it digestible for machine learning algorithms and to extract specific features (Abri et al., 2020). For example, to extract sentiment from a dataset, the review text must be refined in which irrelevant, redundant, noisy, and unreliable information is removed. We then utilize the package NLTK (Natural Language Toolkit) to preprocess the data. NLTK is a standard tool for text analysis used in Python (Berger et al., 2019). We will use the approach that Berger et al. (2019) described to preprocess the data, as depicted in Figure 2:

Figure 3: Preprocessing Flowchart



Firstly, we obtain the relevant variables (Label, Review_Title, and Review_Text) necessary for our data analysis while we drop all the other variables. We make a new variable by contracting the variables 'Review_Title' and 'Review_Text' into the variable 'Review_Text', so all that is written is within one text. As presented in Figure 3, first, we will eliminate any missing values that could potentially corrupt the dataset. Subsequently, the contraction removal process involves expanding literary shortcuts, such as "I've" to "I have," for better readability. Additionally, we remove HTML tags and digits also to increase readability. Then we break down the sentences into a list of word tokens. To reduce variability, each word is converted to its root form. A process that involves tagging each word with its part of speech and then lemmatizing the words into their root forms. The input to the data preprocessing step is the review text. The output of the data preprocessing step is tokens. The tokens serve as input for the lexicons and Linguistic feature extraction.

§ 4.2. Linguistic and Sentiment Analysis Feature Extraction

§ 4.2.1. Linguistic features

As presented in Chapter 2, Theoretical Background, specifically Conceptual Framework, we will use natural language processing code to extract the linguistic features of the review text. Linguistic cues can be extracted from the dataset using various functions. We will extract the following linguistic features from the Amazon dataset: *Quantity*, *Complexity*, and *Diversity*. We will use the package 'collections', specifically 'Counter', to count the frequency of elements within the text. We utilize the 'NLTK' package to extract the part of speech. Besides the 'collections' and 'NLTK', we will use the 'Textstat' package to calculate the Flesch Reading Ease. We will use this function to calculate the readability due to its early success in scientific research. To explore the data, below we present the descriptive statistics of all the extracted linguistic features:

Table 4: Descriptive Statistics Extracted Linguistic Features

Variables	Label	Mean	SD	Min	Max
Number of Words	*Fake*	63.690	58.985	16	1190
	Real	84.212	106.423	16	2851
Number of Sentence	*Fake*	5.435	4.147	1	88
	Real	6.898	6.869	1	198
Number of Caps	*Fake*	10.322	25.399	0	1749
	Real	14.145	28.762	0	1325
Number of Punctuation	*Fake*	10.921	15.748	0	371
	Real	16.553	26.107	0	560
Number of Nouns	*Fake*	16.221	17.078	1	381
	Real	22.677	30.518	1	678
Number of Verbs	*Fake*	11.770	11.049	0	217
	Real	15.047	18.787	0	520
Number of Adjectives	*Fake*	6.155	5.731	0	132
	Real	7.842	10.100	0	261
Number of Adverbs	*Fake*	4.067	3.931	0	71
	Real	4.970	6.575	0	218
Average Word Length	*Fake*	4.376	0.467	3.167	11.922
	Real	4.397	0.433	3.079	8.548
Average Sentence Length	*Fake*	87.290	58.836	18	1038.333
	Real	89.210	63.927	14.375	1465
Redundancy Score	*Fake*	0.540	0.085	0.280	0.955
	Real	0.546	0.087	0.230	0.962
Readability Score	*Fake*	79.476	12.940	-54.39	115.64

	Real	80.328	12.642	-99.56	116.15
Lexical Diversity	*Fake*	0.807	0.098	0.318	1
	Real	0.800	0.111	0.354	1

These features present counts of instances to get numerical information about the content of the review. The descriptive table shows the differences between real and fake reviews. Standing out is the number of words and sentences used in fake vs. real reviews. Looking at Table 4, real reviews are substantially longer than fake reviews. Besides, the average word and sentence length are almost the same for each label. Also, given that fake reviews need to be more persuasive, intuitively, someone would think that there will be more use of capital letters in fake reviews, but as shown by the statistics, the number of caps in the review is much higher for real reviews. Finally, real reviews score higher in every instance except for Lexical Diversity when looking at the mean of the linguistic features.

§ 4.2.2. Sentiment Analysis using Lexicons

In Chapter 3, Method, we concluded that using lexicons SentiWordNet and VADER would be most promising for this research. Because lexicon-based emotion indicators are less susceptible to indirect indicators of sentiment that could produce incorrect sentiment patterns than machine learning methods, like words as 'not', which will shift the direction of the emotion, and lexicons only rely on explicit word-to-sentiment associations (Ahmed et al., 2015).

The lexicon-based approach divides the document into lexemes (basic lexical units) and examines each sentence. Then we classify the words as positive, negative, or neutral opinions based on a pre-defined dictionary. We will use Python to implement this approach. We used the following package to extract the sentiment analysis features: 'nltk.corpus', more specifically 'sentiwordnet', to calculate the sentiment scores and extract individual word sentiment using the SentiWordNet lexicon and 'nltk.sentiment.vader', more specifically 'SentimentIntensityAnalyzer', to calculate the sentiment scores and extract individual word sentiment using the VADER lexicon.

The lexicon-based approach has been reported to achieve better accuracy due to using a pre-defined dictionary (Anees et al., 2020). Given that every lexicon produces different output, we end up with two different sentiment analysis feature sets. The SentiWordNet lexicon will also calculate the number of words within a sentiment category to get more insights into the emotional structure of the reviews. Within VADER, this is not possible due to a different lexicon approach. We present these results in the following descriptive statistical tables:

Table 5: Descriptive Statistics Extracted Sentiment Analysis Features

		SentiWordNet				VADER			
Variables	**Label**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Sentiment Score	*Fake*	1,549	2,014	-9,125	24,125	0,632	0,530	-0,996	1,000
	Real	1,307	2,306	-10,625	40,5	0,626	0,510	-1,000	1,000
Review Intensity	*Fake*	2,113	1,774	0	37,297	-	-	-	-
	Real	2,471	2,7754	0	86,813	-	-	-	-
Review Ambiguity	*Fake*	41,360	37,982	7	733	-	-	-	-
	Real	54,563	68,044	1	1754	-	-	-	-
Number of Positive Words	*Fake*	8,649	7,237	0	142	5,039	3,867	0	61
	Real	9,954	11,756	0	313	5,450	5,292	0	108
Number of Negative Words	*Fake*	3,812	4,114	0	110	1,217	1,991	0	67
	Real	5,247	6,819	0	196	1,628	3,416	0	213
Number of Objective/Neutral Words	*Fake*	61,250	62,065	10	1252	53,424	51,520	11	1052
	Real	84,301	111,811	13	2835	72,154	93,949	11	2433
Positive Score	*Fake*	-	-	-	-	0,226	0,128	0	0,708
	Real	-	-	-	-	0,206	0,123	0	0,672
Negative Score	*Fake*	-	-	-	-	0,047	0,065	0	0,51
	Real	-	-	-	-	0,047	0,059	0	0,519
Neutral Score	*Fake*	-	-	-	-	0,727	0,113	0,292	1
	Real	-	-	-	-	0,748	0,114	0,328	1
Polarity Shifters	*Fake*	-	-	-	-	0,458	0,806	0	14
	Real	-	-	-	-	0,715	1,156	0	15
Intensity Modifiers	*Fake*	-	-	-	-	1,636	1,719	0	25
	Real	-	-	-	-	1,754	2,404	0	71
Negations	*Fake*	-	-	-	-	0,973	1,376	0	18
	Real	-	-	-	-	1,294	1,886	0	47
Emoticons	*Fake*	-	-	-	-	0,019	0,153	0	4
	Real	-	-	-	-	0,026	0,178	0	6

The sentiment analysis features extracted from the Amazon fake review dataset present exciting insights. The sentiment score for both lexicons shows that the mean is positive, indicating that most reviews are written in a positive tense. Looking at the number of appearances of positive, negative, and objective/neutral words, the number of these polarity indicators is higher for real reviews than for fake. Research has suggested that fake reviews express more emotion than real ones (Vaitheeswaran & Arockiam, 2016). The overall sentiment score for both lexicons confirms that hypothesis. However, looking at Table 5, specifically at the number of polarity words, our data suggests otherwise. The mean is higher in all instances for real reviews, and looking at the maximum numbers, they are way higher, suggesting more expression of extreme emotions.

§ 4.3. Final dataset

After the procedure of preprocessing and extracting all the linguistic and sentiment analysis features, two appliable datasets will remain. The first dataset ('senti_df') we based on the SentiWordNet-Lexicon, and the second dataset ('VADER_df") we based on the VADER-Lexicon. The SentiWordNet-Lexicon dataset consists of 36 columns (including the Label of the review) consisting of different linguistic and emotional features. The VADER-Lexicon dataset consists of 25 columns (including the Label of the review) representing the various features. In Appendix 1 the names of the gathered features are presented. Before using the extracted feature within the machine learning classifiers, we will normalize the data to make it operational. In machine learning, normalization is converting data into the range of 0 to 1 (Wang & Sun, 2014). Because normalization ensures that all variables have similar scales and ranges, so the model has no dominant contributions. Normalization makes the numeric outputs comparable and easier for machine learning algorithms to learn. Researchers have proven that normalization will improve the performance of Support Vector Machine and Gaussian Naïve Bayes (Feng & Palomar, 2015; Rezaeian & Novikova, 2020). Finally, we will combine the two lexicon feature sets into one complete set ('sentiVADER_df'). Presented in Appendix 1, including 42 columns (including the Label of the review). We excluded the overlapping features so there are no double presented features that can influence the performance of the detection model. When we completed this step, the dataset was ready to be utilized as input for the fake review detection methods, SVM and NB.

§ 4.4. Comparison of Features

To compare the significance of the features of fake and real reviews we use the independent two-sample t-tests on the two feature sets. This test compares the means of our specific features between two groups (fake vs. real) to assess whether there is a significant difference (Martens & Maalej, 2019). The t-test results are presented in Appendix 2, Table 6 and Table 7.

The t-tests revealed significant differences between fake and real reviews for almost all the investigated features ($p < 0.05$) using both the VADER and SentiWordNet lexicons. These results suggest that the linguistic and sentiment analysis features extracted from both lexicons exhibit substantial variations between fake and real reviews. As shown in the tables all linguistics features are significant, indicating its impact in detecting fake reviews.

Looking at the T-statistics of the linguistic features *Number of words*, *Number of sentences*, and *Number of punctuations*, they demonstrate substantial differences, indicating their crucial role in distinguishing fake reviews from real reviews. All present a negative number, indicating that the mean of these features is lower for fake reviews, and thereby present shorter reviews.

Furthermore, looking at the SentiWordNet lexicon, features such as *Sentiment score, Review intensity*, *Number of positive words*, and *Number of negative words* also exhibited significant differences between fake and real reviews, implying that sentiment polarity and the presence of positive or negative words serve as important cues in identifying fake reviews.

However, it is worth noting that *Sentiment score* and *Negative score* were insignificant when looking at the results of the VADER lexicon. Given the fact that the Sentiment score is significant when using the SentiWordNet lexicon, it might be the case this lexicon is more appropriate for this data and objective than VADER. This suggests that these features may be more relevant or sensitive to particular lexicons, highlighting the importance of lexicon selection in analyzing fake reviews.

In conclusion, both the VADER and SentiWordNet lexicons offer valuable linguistic and sentiment analysis features for the detection of fake reviews. The observed significant differences across various features underscore the potential of these features in enhancing the accuracy and reliability of machine learning algorithms for fake review detection.

§ 4.5. Feature Selection

To investigate the effectiveness of the linguistic and sentiment analysis features (lists of all features in Appendix 1), we apply a feature selection and reduction technique. Since we are unfamiliar with the importance of the features, we utilize the Recursive Feature Elimination (RFE) method to evaluate and select the most promising features and improve the classification performance (Chen & Jeong, 2007).

Random Forest Recursive Feature Elimination (RF-RFE) is a technique we use to select the most important and best features for the implementation of our Support Vector Machines (SVM) and Naive Bayes (NB) algorithms. Here's how Random Forest RFE works in the context of these machine-learning models:

First, the Random Forest model is trained on the entire feature set (Appendix 1). Random Forest is an ensemble learning method that combines multiple decision trees to make predictions (Abri et al., 2019). We chose the Random Forest algorithm because it is capable of ranking features based on their importance. The Random Forest model we train assigns

importance scores to each feature based on how much they contribute to the overall accuracy of the model. The importance scores are calculated by measuring the decrease in accuracy when a particular feature is randomly assigned within the feature list or removed from it. The features are ranked according to their importance scores in descending order (Misra & Yadav, 2020). The higher the score, the more important the feature is. Starting from the feature with the lowest importance score, features are progressively eliminated one by one by the RFE model, and the model is retrained each time without the eliminated feature. This process continues until one feature remains to evaluate the performance (Martens & Maalej, 2019).

By iteratively eliminating the least important features, we obtain a feature list ordered by their significance for the detection task. The generated feature list from RF-RFE serves as the input for our SVM and NB classifiers. As we aim to identify the best number and group of features for optimal performance, we follow a systematic process of reducing the input features, one by one, in each iteration. To sum up, we train the SVM and NB models using the feature list obtained from RF-RFE, then evaluate the metrics, identify the feature with the lowest importance score and remove it, and repeat the process with an updated list until one feature is remaining.

By using Random Forest RFE, we can identify the most relevant linguistic and sentiment analysis features for detecting fake reviews using SVM and NB. This approach helps to understand which features are essential for achieving accurate predictions, and results in a list with the least to most important features.

Chapter 5: Results

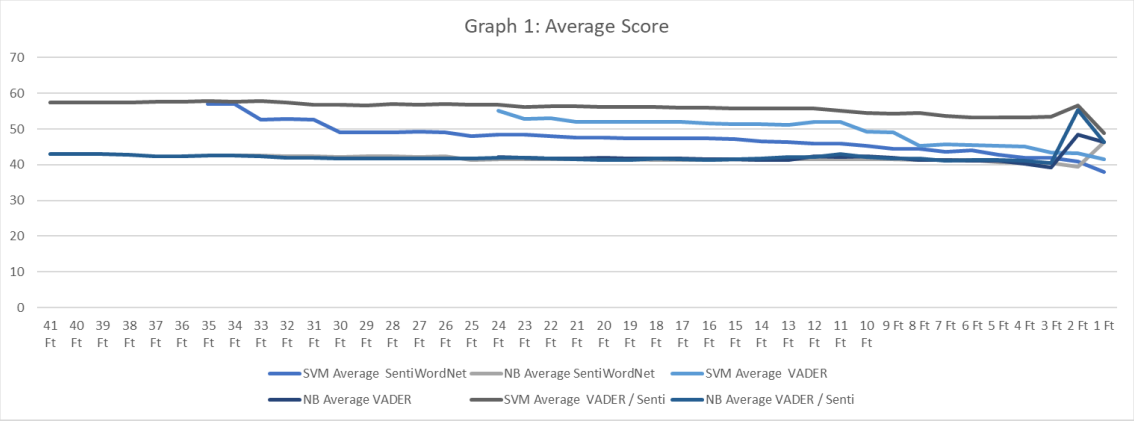
This research aims to find to what extent a machine learning algorithm can detect fake reviews using only linguistic and sentiment analysis features in e-commerce and which linguistic and sentiment analysis features are most effective. We will first address the first part of the research question, and in the second section, we will present the results regarding the second part of the question. As stated, this research will not use the text of the review but rather extract all the linguistic and sentiment features to train the classifier to detect fake or genuine reviews. This way, we can truly research the effectiveness of these features in detecting fake reviews.

In order to distinguish between fake and real online reviews, the analysis of transcribed textual features is the main emphasis of this thesis. Most of the time when text is analyzed, the focus is on linguistic features. We will add sentiment analysis features to add an extra layer to understand the way the review is written. Abri et al. (2020) already used only linguistic features to detect fake reviews but did not include sentiment analysis features. We use both linguistic and sentiment analysis features to find a new feature set to better understand fake reviews and to be able to build a better detection method.

We used natural language processing to extract the linguistic and sentiment analysis features of each of the 21,000 reviews and looked at two promising sentiment lexicons: VADER and SentiWordNet. Numerous numerical features (VADER: 24 & SentiWordNet: 35) were extracted from the text to function as 'food' for the machine learning detection algorithm. We will evaluate the feature lists of VADER, SentiWordNet, and both lexicons combined. We will evaluate which features work best in which list to fully understand the working of these features. To result in a complete list when using both lexicons, we deleted any overlapping variables, e.g., the number of positive words. Due to the fact two sentiment scores would be useless, we will not use the sentiment score from the VADER lexicon because as the t-test showed, it is not a significant factor, but the SentiWordNet score. The total set consists of 41 features. The three sets will be used as input for the SVM and NB algorithm, we split the data into an 80/20 distribution, resulting in a training set (16,800 reviews) and a test set (4,200 reviews). All metrics outputs, as described in Chapter 3, to evaluate the performance of the features and detection methods can be consulted in appendices 1 to 5.

§ 5.1. Model Fit

Because we removed the text as a variable, we used the RFE to select the most promising feature set, eliminating one at a time. As shown in Graph 1, the lines start at different points because this was the maximum number of features extracted using the lexicons and the combination of the lexicons into one dataset. Graph 1 displays the average score calculated by summing the metrics Accuracy, Precision, Recall, and F1-score, and dividing it by four. The Ft's on the X-axes are the number of features extracted by RFE to compare its performance.

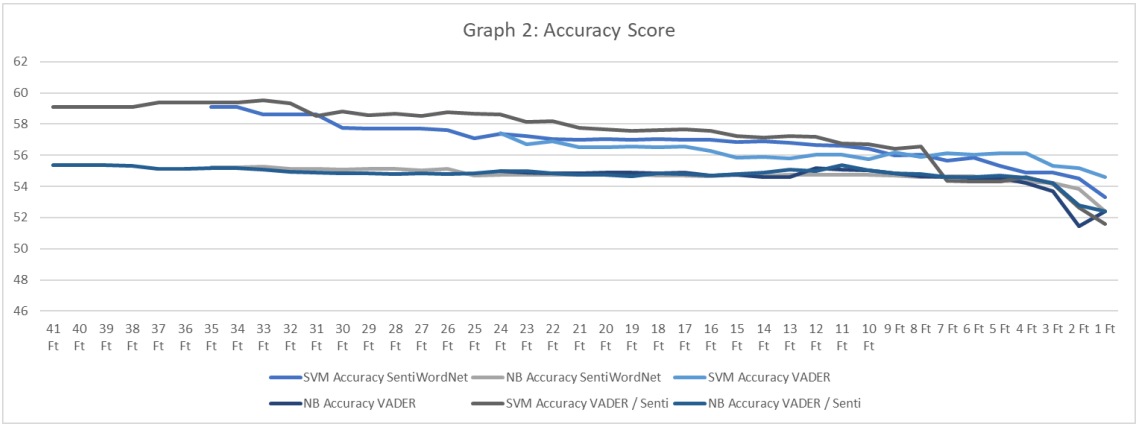


We plotted the metrics derived from the feature sets that were created based on the Amazon fake review dataset. The patterns show the distribution of the performance of the three-dataset using SVM and NB. Looking at the results (Appendix 3), we can conclude that SVM performs at a higher level than NB. To answer to what extent we can detect fake reviews based on only textual features, we look at the highest average score. When we combined the VADER and SentiWordNet lexicons, we achieved an average score of 57,75 using 35 features, presented in Appendix 8.

To further research the influence of only linguistic and sentiment analysis features on fake review detection, we will discuss the metrics accuracy, precision, recall, and F1-score, as presented in Chapter 3, Model.

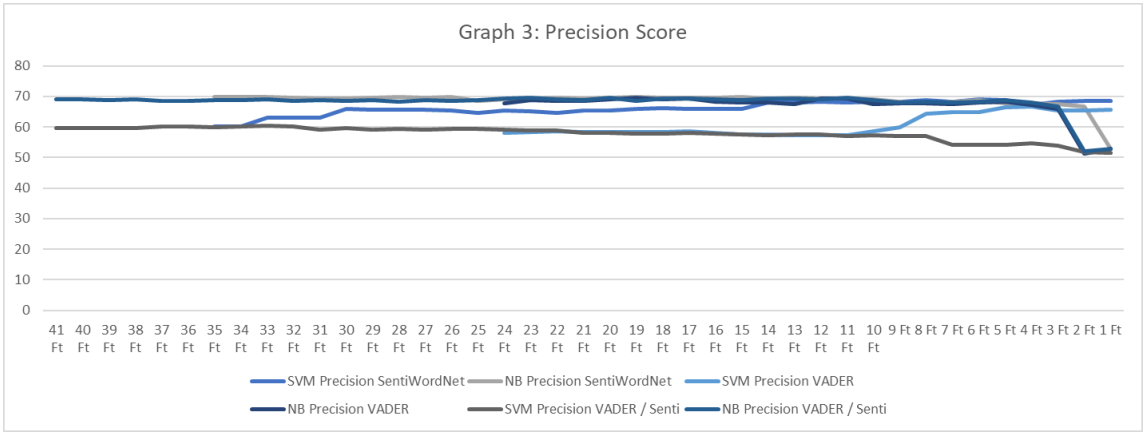
§ 5.1.1. Accuracy

First of all, accuracy. We present the output of the detection method in Appendix 4. The highest Accuracy-score of 60% was achieved by the SVM algorithm using 33 Features from both the VADER and SentiWordNet lexicon. The feature list is presented in Appendix 9.
This is, as the literature showed, a low accuracy score. In the graph below (Graph 2) we can spot a decreasing line from 41 selected features to one. Again, NB is underperforming compared to SVM using all of the three different feature lists.

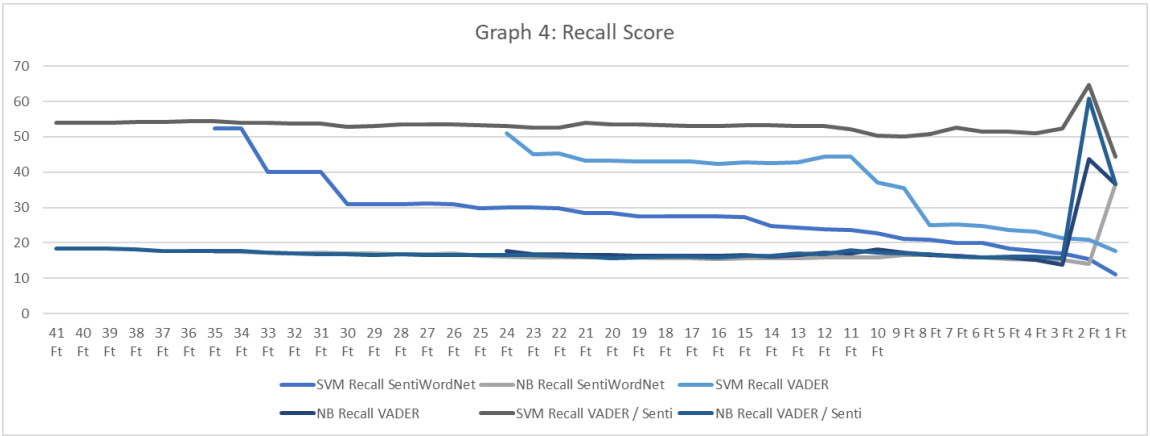


§ 5.1.2. Precision and Recall

The Highest precision score was presented by the Naïve Bayes algorithm with 70% (Appendix 5) for both SentiWordNet (using 28 features), VADER (using 19 features), and combined (using 23 features). The models' precision did not fluctuate much, as displayed in Graph 3.



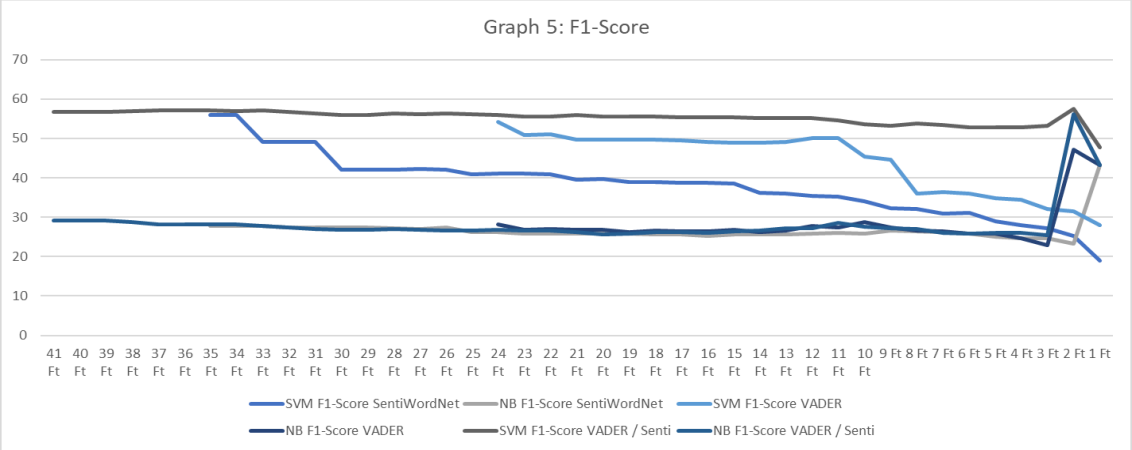
Besides, the NB algorithm performed poorly when looking at the Recall-score in Graph 4, where SVM had a maximum score of 65% using 2 features (Appendix 6).



§ 5.1.3. F1-Score

Finally, the NB algorithm performed at a low level when evaluating the F1-score. It provides a single statistic showing the weighted average of precision and recall. The results of the F1 measure based on all the features and methods are presented in Appendix 7.

The SVM-combined-lexicons approach achieved the highest score, with 58 using only 2 features. When analyzing the results and reviewing the graph of the SVM-combined-lexicons model, the F1-score doesn't fluctuate much. So, when it comes to F1-score it doesn't make a big difference how many features there are used.



§ 5.1.4. Interim Conclusion

When looking at the differences in scores in the performance evaluation tables in Appendices 1 to 5, we conclude that the SVM algorithm performed better than the NB algorithm. The highest overall score was achieved using the SVM algorithm in combination with all 35 features (Appendix 8) from the combined lexicon feature list. With an accuracy of 59%, it is not performing at its best. This shows that the usage of text to detect fraudulent reviews is a must, but not the most important variable. More in-depth research is needed to achieve higher performance using Linguistic and Sentimental features to detect fake reviews, but this research shows the potential that can't be overlooked, due to the fact it was still possible to distinguish fake from real reviews.

§ 5.2. Feature Importance

To address the second part of our research question we will use feature importance. To find out which features are most effective. Because we utilized two different lexicons, SentiWordNet and VADER, this results in two distinct feature sets. We then combined these feature sets, excluding overlapping features, to generate a third extensive feature set. The sets are used for the RFE selector's Random Forest algorithm, to evaluate each feature's importance. The list of features and their importance scores are presented in Tables 8, 9 and 10.

The descending order of feature importance provides a clear hierarchy, enabling researchers, professionals, and consumers to get insights into the underlying mechanisms and dynamics of fake review detection. This prioritization allows for a more targeted classification approach, focusing on the most influential features when developing a detection model.

The presented lists of features provide specific insights into the linguistic and sentiment analysis features that are most relevant for the detection of fake reviews. The VADER Lexicon list, for instance, highlights the importance of sentiment analysis features related to positive score, neutral score, and sentiment ratings. The SentiWordNet Lexicon list, on the other hand, emphasizes the importance of elements like the number of nouns and lexical diversity. Finally, in the combined feature list we can conclude that both linguistic and sentiment feature are important indicators to evaluate the nature of a review. These lists present linguistic and sentiment analysis features, enabling the development of models for fake review identification that are more effective and accurate.

Table 8: Importance VADER

VADER Feature Importance		
Rank	Feature	Importance
24	*emoticons*	0,0024
23	*negative words*	0,0107
22	*negations*	0,0130
21	*positive words*	0,0156
20	*neutral words*	0,0241
19	*polarity shifters*	0,0249
18	*Number of adverbs*	0,0253
17	*intensity modifiers*	0,0263
16	*Number of adjectives*	0,0264
15	*negative score*	0,0268
14	*Number of verbs*	0,0289
13	*Number of sentences*	0,0299
12	*Number of caps*	0,0407
11	*Number of nouns*	0,0416
10	*Lexical diversity*	0,0520
9	*Average sentence length*	0,0521
8	*positive score*	0,0526
7	*Average word length*	0,0534
6	*neutral score*	0,0542
5	*Readability score*	0,0620
4	*Number of words*	0,0664
3	*sentiment score*	0,0705
2	*Redundancy score*	0,0752
1	*Number of punctuation*	0,1249

Table 9: Importance SentiWordNet

SentiWordNet Feature Importance		
Rank	Feature	Importance
35	*pos 1*	0,0000
34	*neg 1*	0,0001
33	*neg 875*	0,0033
32	*pos 75*	0,0046
31	*pos 875*	0,0057
30	*neg 75*	0,0059
29	*neg 375*	0,0078
28	*pos 375*	0,0098
27	*neg 625*	0,0102
26	*neg 5*	0,0102
25	*pos 25*	0,0117
24	*neg 125*	0,0125
23	*neg 25*	0,0130
22	*pos 625*	0,0136
21	*pos 5*	0,0142
20	*pos 125*	0,0153
19	*Number of negative words*	0,0207
18	*Number of verbs*	0,0217
17	*Number of positive words*	0,0217
16	*Number of adjectives*	0,0218
15	*Number of adverbs*	0,0244
14	*Number of sentences*	0,0277
13	*Review ambiguity*	0,0286
12	*Number of objective words*	0,0310
11	*Number of caps*	0,0401
10	*Number of nouns*	0,0409
9	*Average sentence length*	0,0486
8	*Review intensity*	0,0504
7	*Average word length*	0,0505
6	*Lexical diversity*	0,0517
5	*Readability score*	0,0560
4	*Number of words*	0,0649
3	*Sentiment score*	0,0687
2	*Redundancy score*	0,0711
1	*Number of punctuation*	0,1216

Table 10: Importance Combined

VADER / Senti Feature Importance		
Rank	Feature	Importance
41	*pos 1*	0,0000
40	*neg 1*	0,0001
39	emoticons	0,0016
38	*neg 875*	0,0026
37	*pos 75*	0,0043
36	*neg 75*	0,0049
35	*pos 875*	0,0054
34	*neg 375*	0,0062
33	*negations*	0,0072
32	*neg 625*	0,0076
31	*neg 5*	0,0080
30	*pos 375*	0,0098
29	*pos 625*	0,0105
28	*neg 25*	0,0105
27	*pos 25*	0,0107
26	*neg 125*	0,0111
25	*pos 5*	0,0116
24	*pos 125*	0,0125
23	*Number of negative words*	0,0144
22	*Number of positive words*	0,0174
21	*Number of adjectives*	0,0185
20	*Number of adverbs*	0,0189
19	intensity modifiers	0,0189
18	*Number of verbs*	0,0194
17	polarity shifters	0,0199
16	*Number of objective words*	0,0230
15	*Review ambiguity*	0,0239
14	*Number of sentences*	0,0241
13	*Number of caps*	0,0344
12	*Number of nouns*	0,0360
11	*Average word length*	0,0405
10	*Lexical diversity*	0,0424
9	*Average sentence length*	0,0432
8	*Review intensity*	0,0435
7	*Sentiment score*	0,0464
6	*Readability score*	0,0502
5	neutral score	0,0506
4	Number of words	0,0567
3	Redundancy score	0,0601
2	positive score	0,0638
1	*Number of punctuation*	0,1092

Chapter 6: Discussion

§ 6.1. Summary of Main Findings

As described in Chapter 1, Introduction, our main contributions will be to illustrate how we will use textual features to predict and understand fake reviews without the actual text. First, we extracted the sentiment analysis features sentiment score, positivity, negativity, and objectivity, number of words in different sentiment categories (positive, negative, objective/neutral), and sentiment score per word by applying two different lexicons (VADER vs. SentiWordNet). This resulted in three feature sets including quantity, complexity, diversity, and sentiment. The feature sets (24 features VADER, 35 features SentiWordNet, and 41 features combined lexicons) all had a different order of importance within the list of features. Looking at the lists, we conclude that the Readability score, Number of words, Sentiment score, Redundancy score, and Number of punctuations are important features. As shown in the top 10 features (tables 7,8, & 9), both linguistic and sentiment analysis features are present, confirming the hypothesis of Berger et al. (2019) that both classes of features are important indicators to detect a fake review.

We conclude that VADER highlights the importance of sentiment analysis features related to positive scores, neutral scores, and sentiment ratings. Furthermore, the SentiWordNet Lexicon emphasizes the importance of elements like review intensity and lexical diversity. Analyzing the combined list, we can see that the top 10 is a combination of the top 10's of the separate lexicon feature lists. The combined list showed better and more steady results in comparison to the separate lexicon lists, which makes it more promising for future research.

We also conclude that SentiWordNet is a slightly better lexicon for classifying fake reviews in combination with machine learning (56.92 > 55.17), also because VADER showed insignificant features like sentiment score and negative score within the feature set. Our results showed that SVM performs better than NB. Finally, the list of features we produced enables future research to develop more effective and accurate models and transparency of how to detect fake reviews for researchers, businesses, and consumers.

§ 6.2. Theoretical and Managerial Takeaways

Our main contribution to the existing literature is combining a lexicon approach with a machine learning approach for an e-commerce dataset based on only textual information. The studies of Abri et al. (2020), Vanta & Aono (2019), and Dewang & Singh (2015) all used different kinds of linguistic and sentiment analysis features to detect fake reviews. They

identified several features as significant predictors. Abri et al. (2020) identified the number of adjectives, redundancy, and lexical diversity as important features, Vanta and Aono (2019) concluded that word count is a significant indicator; Dewang and Singh (2015), and Azimi et al. (2022) connected POS count and polarity score to detect fake reviews. Lastly, Ghose and Ipeirotis (2011) detected readability as an important predictor. Looking at our proposed feature set, all these features play a significant role in identifying fake reviews. We combined these features, including more, into one set of variables to research their impact. This set of features will give research handles for researching and developing fake review detection.

To compare the performance of our model, we will sum up the result of research that used similar methods. Abri et al. (2020) used multiple linguistic features, including the text itself, in combination with Recursive Feature Elimination to achieve an accuracy score of 77% using SVM and 84% using NB. Vanta & Aono (2019) used linguistic, lexicon-extracted features, rating, and text to detect fake reviews. They achieved an accuracy score of 77% using SVM. On average the percentage of accurate prediction by detection models build with SVM and NB was 80% (Choi et al., 2015). In almost every instance SVM outperformed NB. All mentioned papers used besides extracted features from the text other features to improve their model. Most used rating or the text. Our model resulted in an accuracy score of 60%, which is low in comparison to previous studies.

Also, research that used lexicons in combination with Machine learning resulted in better performances than our model. Peng & Zhong (2014) achieved an accuracy of 61.4% using the SentiWordNet lexicon, Wang et al. (2020) using the SentiNet lexicon resulted in 69.9% accuracy using NB and 78.9 using SVM, and finally Taqiuddin et al. (2021) used the TextBlob lexicon in combination with SVM for a result of 81% accuracy. As shown, our models' accuracy is low in comparison to previous studies. We can conclude that only using linguistic and sentiment analysis features is not sufficient to classify fake reviews. The model is lacking other important factors, e.g., rating, product meta-data, or user behavior. So, what we can learn from our results is that there are important elements that have to be taken into account when building a fake review detection model. We present a feature list with many different features for future research, but as shown by the poor performance of the model, detecting fake reviews is complex and tricky, so more factors have to be included.

Besides, these features will provide more clarity and transparency on how to spot fake reviews. Studies have shown that consumers need help distinguishing fake from genuine reviews, and our feature set can be a solution (Azimi et al., 2022). In addition, textual information like the extracted features is easier for consumers to spot and create a more critical

mindset toward the distribution of fake reviews (Hu et al., 2011). Currently, Amazon has no transparency on how they combat fake reviews or how consumers can detect them based on guidelines (Hill, S. 2022). Furthermore, the feature set gives companies, besides Amazon, helpful information to combat fake reviews because preventing fake reviews from appearing on websites is better than curing the problem when it is there.

In short, thanks to our research, there is a feature set available for researchers to investigate fake reviews further and help consumers and businesses understand the detection of these reviews better to ensure no harm is done in the future.

§ 6.3. Limitations and Future Research

Unfortunately, the performance of our machine learning algorithms was not as high as expected, looking at prior fake review detection research. Fake reviews can be very sophisticated and challenging to detect, especially if they are crafted to appear genuine. Relying solely on sentiment and linguistic features might not be sufficient to capture these complex patterns. The model we used might not be complex enough to capture the nuances of the data. More sophisticated models like deep learning methods could potentially improve performance.

Due to this, it is difficult to conclude how predictive our feature set is. Furthermore, we only focused on two sentiment lexicons and machine learning algorithms when there are multiple different methods to apply. Moreover, we used the Amazon fake review dataset. Besides, it is used in multiple scientific papers, there is no clear documentation of the construction and labeling of the dataset. Lastly, we used an e-commerce dataset which limits the scope to only this sector. Fake reviews are a problem across multiple fields, so it is hard to say if the feature set we produced applies to other industries.

The model still needs some fine-tuning to improve the accuracy score, but we could not perform this due to time pressure. In the future, we would like to use real-time datasets across different industries to compare the feature set. A more generalizable real-time dataset will give new insights into building a detection model. We would also like to use different machine (deep)learning algorithms like Multi-Layer Perceptron (MLP). These machine learning models are neural networks, and some literature has already suggested that they achieve more excellent performance than ‘old-school’ models like SVM and NB. We already ran a test using MLP, and the result looks promising. Finally, we would like to add more features to our set, like ratios and reviewer-based features, and evaluate the performance of the model when including other factors. Our detection method did not achieve its full potential, and these features will potentially boost the performance of the detection model.

Chapter 7: Reference list

Aayush. (2018). *GitHub - aayush210789/Deception-Detection-on-Amazon-reviews-dataset*. GitHub. https://github.com/aayush210789/Deception-Detection-on-Amazon-reviews-dataset

Abri, F., Gutiérrez, L. F., Namin, A. S., Jones, K. S., & Sears, D. R. (2020, December). Linguistic features for detecting fake reviews. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 352-359). IEEE.

Aesuli. (2022). *GitHub - aesuli/SentiWordNet: The SentiWordNet sentiment lexicon*. GitHub. https://github.com/aesuli/SentiWordNet

Alsubari, S. N., Shelke, M. B., & Deshmukh, S. N. (2020). Fake reviews identification based on deep computational linguistic. *International Journal of Advanced Science and Technology*, *29*(8s), 3846-3856.

Aono, T. V. M. (2019). Fake review detection focusing on emotional expressions and extreme rating. *The association for natural language processing*.

Azimi, S., Chan, K., & Krasnikov, A. (2022). How fakes make it through: the role of review features versus consumer characteristics. *Journal of Consumer Marketing*, (ahead-of-print).

Baishya, D., Deka, J. J., Dey, G., & Singh, P. K. (2021). SAFER: sentiment analysis-based fake review detection in e-commerce using deep learning. *SN Computer Science, 2*, 1-12.

Bajaj, A. (2023). Can Python understand human feelings through words? – A brief intro to NLP and VADER Sentiment Analysis. *Analytics Vidhya.* https://www.analyticsvidhya.com/blog/2021/06/vader-for-sentiment-analysis/

Bandhakavi, A., Wiratunga, N., Padmanabhan, D., & Massie, S. (2017). Lexicon based feature extraction for emotion text classification. *Pattern recognition letters, 93*, 133-142.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of marketing*, *84*(1), 1-25.

Bharatkumar, J., Kartik, M., Shetty, K., Shreyas Pai, K., & Kumar, S. (2022). E-Commerce Site's Fake Review Detection and Sentiment Analysis using ML Technique. *International Journal of Advanced Research in Computer and Communication Engineering, 11(7),* 180–182. https://doi.org/10.17148/IJARCCE.2022.11733

Birim, Ş. Ö., Kazancoglu, I., Mangla, S. K., Kahraman, A., Kumar, S., & Kazancoglu, Y. (2022). Detecting fake reviews through topic modelling. *Journal of Business Research, 149*, 884-900

Borah, A., Banerjee, S., Lin, Y. T., Jain, A., & Eisingerich, A. B. (2020). Improvised marketing interventions in social media. *Journal of Marketing*, *84*(2), 69-91.

Calderon, P. (2018, March 31). VADER Sentiment Analysis Explained - Pio Calderon - Medium. *Medium*. https://medium.com/@piocalderon/vader-sentiment-analysis-explained-f1c4f9101cd9

Chen, X. W., & Jeong, J. C. (2007, December). Enhanced recursive feature elimination. In *Sixth international conference on machine learning and applications (ICMLA 2007)* (pp. 429-435). IEEE.

Cheng, L. C., Tseng, J. C., & Chung, T. Y. (2017, July). Case study of fake web reviews. In *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017* (pp. 706-709).

Choi, W., Nam, K., Park, M., Yang, S., Hwang, S., & Oh, H. (2022). Fake review identification and utility evaluation model using machine learning. *Frontiers in artificial intelligence*, *5*.

Cjhutto. (2022). GitHub - *cjhutto/vaderSentiment: VADER Sentiment Analysis*. GitHub. https://github.com/cjhutto/vaderSentiment

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415-443.

Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*, *34*(6), 480-488.

Duwairi, R. M., Ahmed, N. A., & Al-Rifai, S. Y. (2015). Detecting sentiment embedded in Arabic social media–a lexicon-based approach. *Journal of Intelligent & Fuzzy Systems*, *29*(1),107-117.

Elmogy, A. M., Tariq, U., Ammar, M., & Ibrahim, A. (2021). Fake reviews detection using supervised machine learning. *International Journal of Advanced Computer Science and Applications*, *12*(1).

Elmurngi, E., & Gherbi, A. (2017). Detecting fake reviews through sentiment analysis using machine learning techniques. *DATA ANALYTICS*, 9.

FATTAHI, S., OTHMAN, Z., & OTHMAN, Z. A. (2015). NEW APPROACH FOR IMBALANCED BIOLOGICAL DATASET CLASSIFICATION. *Journal of Theoretical & Applied Information Technology*, *72*(1).

Feng, Y., & Palomar, D. P. (2015). Normalization of linear support vector machines. *IEEE Transactions on Signal Processing*, *63*(17), 4673-4688.

Garcia, L. (2020, September 25). Deception on Amazon — an NLP exploration — Part 1 - Liev Garcia - Medium. *Medium*. https://medium.com/@lievgarcia/deception-on-amazon-c1e30d977cfd

Gavilan, D., Avello, M., & Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, *66*, 53-61.

Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE transactions on knowledge and data engineering*, *23*(10), 1498-1512.

Gupta, S., & Mandal, S. (2017). Supervised machine learning vs. Lexicon-based text classification for sentiment analysis: A comparative study. In *Computer, Communication and Electrical Technology* (pp. 55-59). CRC Press.

Hassan, R., & Islam, M. R. (2021, February). Impact of sentiment analysis in fake online review detection. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)* (pp. 21-24). IEEE.

He, S., Hollenbeck, B., Overgoor, G., Proserpio, D., & Tosyali, A. (2022). Detecting fake-review buyers using network structure: Direct evidence from Amazon. *Proceedings of the National Academy of Sciences*, *119*(47), e2211932119.

He, X. (2023). *Essays on Platform Policies, Ratings and Innovation* (Doctoral dissertation, UCLA).

Hill, S. (2022, August 22). How to Spot Fake Reviews on Amazon (2023): Tools and Advice. WIRED. https://www.wired.com/story/how-to-spot-fake-reviews-amazon/

Hossain, M. F. (2019). Fake review detection using data mining.

How to create a fake review detection model. (2022, September 6). https://practicaldatascience.co.uk/machine-learning/how-to-build-a-fake-review-detection-model#:~:text=The%20review%20text%20is%20the,review%20is%20fake%20or%20not.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177).

Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision support systems*, *52*(3), 674-684.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013, May). Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 607-618).

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Jain, K. (2021, December 31). How to Improve Support Vector Machine? Medium. https://kopaljain95.medium.com/how-to-improve-support-vector-machine-9561ab96ed18

Jeff Bezos Quotes. (1997). BrainyQuote. https://www.brainyquote.com/quotes/jeff_bezos_173310

Karami, A., & Zhou, B. (2015). Online review spam detection by new linguistic features. *iConference 2015 Proceedings*.

Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, *90*, 523-537.

Kntb. (2021). *GitHub - kntb0107/fake_review_detector: Project Files for Final Year Project.* GitHub. https://github.com/kntb0107/fake_review_detector/tree/main

Krishna, A., Akhilesh, V., Aich, A., & Hegde, C. (2019). Sentiment analysis of restaurant reviews using machine learning techniques. In *Emerging Research in Electronics, Computer Science and Technology: Proceedings of International Conference, ICERECT 2018* (pp. 687-696). Springer Singapore.

Krishnamoorthy, S. (2015). Linguistic features for review helpfulness prediction. *Expert Systems with Applications*, *42*(7), 3751-3759.

Kumar, R., Mukherjee, S., & Rana, N. P. (2023). Exploring Latent Characteristics of Fake Reviews and Their Intermediary Role in Persuading Buying Decisions. *Information Systems Frontiers*, 1-18.

Lai, C. L., Xu, K. Q., Lau, R. Y., Li, Y., & Jing, L. (2010, November). Toward a language modeling approach for consumer review spam detection. In *2010 ieee 7th international conference on e-business engineering* (pp. 1-8). IEEE.

Lievcin. (2021). *GitHub - lievcin/amazon_deception: NLP project analysing the Amazon Product reviews dataset*. GitHub. https://github.com/lievcin/amazon_deception

Louiefb. (2019). *GitHub -louiefb/amazon-reviews-nlp.* GitHub. https://github.com/louiefb/amazon-reviews-nlp/blob/master/Amazon%20Reviews%20NLP.ipynb

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, *62*(12), 3412-3427.

Machova, K., Mach, M., & Vasilko, M. (2022). Comparison of machine learning and sentiment analysis in detection of suspicious online reviewers on different type of data. *Sensors*, *22*(1), 155.

Mahmood, A., Kamaruddin, S., Naser, R., & Nadzir, M. (2020). A combination of lexicon and machine learning approaches for sentiment analysis on Facebook. *J. Syst. Manag. Sci*, *10*(3), 140-150.

Martens, D., & Maalej, W. (2019). Towards understanding and detecting fake reviews in app stores. *Empirical Software Engineering*, *24*(6), 3316-3355.

McCluskey, M. (2022, July 6). Inside the War on Fake Consumer Reviews. *Time*. https://time.com/6192933/fake-reviews-regulation/

Misra, P., & Yadav, A. S. (2020). Improving the classification accuracy using recursive feature elimination with cross-validation. *Int. J. Emerg. Technol*, *11*(3), 659-665.

Mitra, A. (2020). Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset). *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, *2*(03), 145-152.

Moqueem, A., Moqueem, F., Reddy, C. V., Jayanth, D., & Brahma, B. (2023, January). Online Shopping Fake Reviews Detection Using Machine Learning. In *Cognition and Recognition: 8th International Conference, ICCR 2021, Mandya, India, December 30–31, 2021, Revised Selected Papers* (pp. 305-318). Cham: Springer Nature Switzerland.

Mueller, K, Fakespot(2023, February 22). Amazon, Walmart, eBay, other sellers, targeted with scam that could compromise your accounts. *Nj*. https://www.nj.com/news/2022/12/amazon-walmart-ebay-other-sellers-targeted-with-scam-that-could-compromise-your-accounts.html

Nafis, N. S. M., & Awang, S. (2021). An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access*, *9*, 52177-52192.

Narayana Royal, M., Reddy, R. P. K., Sangathya, G. S., Sai Madesh Pretam, B., Kaliappan, J., & Suganthan, C. (2023). Detection of Fake Reviews on Products Using Machine Learning. In *Information and Communication Technology for Competitive Strategies (ICTCS 2022)* (pp. 601-611). Springer, Singapore.

Nguyen, H., Veluchamy, A., Diop, M., & Iqbal, R. (2018). Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. *SMU Data Science Review*, *1*(4), 7.

NLTK :: nltk.sentiment.vader. (n.d.). https://www.nltk.org/_modules/nltk/sentiment/vader.html

Online consumer reviews: The case of misleading or fake reviews | Think Tank | European Parliament. (2015). https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2015)571301

Otero, J. M. M. (2021). Fake reviews on online platforms: perspectives from the US, UK and EU legislations. *SN Social Sciences*, *1*(7).

Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism management*, *32*(2), 215-224.

Patel, D., Kapoor, A., & Sonawane, S. (2018). Fake review detection using opinion mining. *International Research Journal of Engineering and Technology (IRJET)*, *5*.

Pendyala, A. (2019). *Fake consumer review detection* (Doctoral dissertation, California State University, Sacramento).

Peng, Q., & Zhong, M. (2014). Detecting Spam Review through Sentiment Analysis. *J. Softw.*, *9*(8), 2065-2072.

Poonguzhali, R., Sowmiya, S. F., Surendar, P., & Vasikaran, M. (2022, April). Fake Reviews Detection using Support Vector Machine. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 1509-1512). IEEE.

Rezaeian, N., & Novikova, G. (2020). Persian text classification using naive bayes algorithms and support vector machine algorithm. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *8*(1), 178-188.

Roul, A. (2021, December 26). Sentiment Analysis- Lexicon Models vs Machine Learning. *Medium*. https://medium.com/nerd-for-tech/sentiment-analysis-lexicon-models-vs-machine-learning-b6e3af8fe746

Salminen, J., Kandpal, C., Kamel, A. M., Jung, S. G., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services*, *64*, 102771.

Sanieldalib. (2019). *GitHub - sanieldalib/Amazon-Review-Classifier: This is a R-Shiny app which classifies Amazon reviews as real or fake simply by pasting the product URL*. GitHub. https://github.com/sanieldalib/Amazon-Review-Classifier

Saumya, S., & Singh, J. P. (2018). Detection of spam reviews: a sentiment analysis approach. *Csi Transactions on ICT*, *6*(2), 137-148.

Shahariar, G. M., Biswas, S., Omar, F., Shah, F. M., & Hassan, S. B. (2019, October). Spam review detection using deep learning. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 0027-0033). IEEE.

Srivastava, R., Bharti, P. K., & Verma, P. (2022). Comparative Analysis of Lexicon and Machine Learning Approach for Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, *13*(3).

Taqiuddin, R., Bachtiar, F. A., & Purnomo, W. (2021). Opinion spam classification on steam review using support vector machine with lexicon-based features. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*.

Vaitheeswaran, G., & Arockiam, L. (2016). Combining lexicon and machine learning method to enhance the accuracy of sentiment analysis on big data. *International Journal of Computer Science and Information Technologies*, *7*(1), 306-311.

Wang, G., Shang, G., Pu, P., Li, X., & Peng, H. (2022). Fake Review Identification Methods Based on Multidimensional Feature Engineering. *Mobile Information Systems*, *2022*.

Wang, J., Kan, H., Meng, F., Mu, Q., Shi, G., & Xiao, X. (2020). Fake review detection based on multiple feature fusion and rolling collaborative training. *IEEE Access*, *8*, 182625-182639.

What is Natural Language Processing? | IBM. (n.d.). https://www.ibm.com/topics/natural-language-processing

Wu, Y., Ngai, E. W., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, *132*, 113280.

Xhymshiti, M. (2020). *Domain independence of Machine Learning and lexicon based methods in sentiment analysis* (Bachelor's thesis, University of Twente).

Xhymshiti, M. (2020). *Domain independence of Machine Learning and lexicon based methods in sentiment analysis* (Bachelor's thesis, University of Twente).

Ye, J., & Akoglu, L. (2015). Discovering opinion spammer groups by network footprints. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I 15* (pp. 267-282). Springer International Publishing.

Yousif, A., & Buckley, J. (2022). Impact of Sentiment Analysis in Fake Review Detection. *arXiv preprint arXiv:2212.08995*.

Yrnigam. (2020). *GitHub - yrnigam/Amazon-Review-for-Sentiment-Analysis-ML:* GitHub. https://github.com/yrnigam/Amazon-Review-for-Sentiment-Analysis-ML

Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, *33*(2), 456-481.

Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, *33*(2), 456-481.

Zhang, H., Gan, W., & Jiang, B. (2014, September). Machine learning and lexicon based methods for sentiment classification: A survey. In *2014 11th web information system and application conference* (pp. 262)

El-Din, D. M. (2016). Enhancement bag-of-words model for solving the challenges of sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *7*(1).

Almestekawy, A., & Abdulsalam, M. (2019). Sentiment analysis of product reviews using bag of words and bag of concepts. *International Journal of Electronics and Information Engineering*, *11*(2), 49-60.

Hamouda, A., Marei, M., & Rohaim, M. (2011). Building machine learning based senti-word lexicon for sentiment analysis. *Journal of advances in information technology*, *2*(4), 199-203.

Augustyniak, L., Kajdanowicz, T., Szymański, P., Tuligłowicz, W., Kazienko, P., Alhajj, R., & Szymanski, B. (2014, August). Simpler is better? Lexicon-based ensemble sentiment classification beats supervised methods. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)* (pp. 924-929). IEEE.

Hajek, P., Barushka, A., & Munk, M. (2020). Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. *Neural Computing and Applications*, *32*, 17259-17274.

Poonguzhali, R., Sowmiya, S. F., Surendar, P., & Vasikaran, M. (2022, April). Fake Reviews Detection using Support Vector Machine. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 1509-1512). IEEE.

Taqiuddin, R., Bachtiar, F. A., & Purnomo, W. (2021). Opinion spam classification on steam review using support vector machine with lexicon-based features. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*.

Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. *Encyclopedia of machine learning*, *15*(1), 713-714.

Chapter 8: Appendices

Appendix 1:

	Features SentiWordNet
1	Number_of_word
2	Number_of_sentences
3	Number_of_caps
4	Number_of_punctuation
5	Number_of_nouns
6	Number_of_verbs
7	Number_of_adjectives
8	Number_of_adverbs
9	Average_word_length
10	Average_sentence_length
11	Redundancy_score
12	Readability_score
13	Lexical_diversity
14	Sentiment_score
15	Review_intensity
16	Review_ambiguity
17	Number_of_positive words
18	Number_of_negative words
19	Number_of_objective words
20	pos_125
21	neg_125
22	pos_25
23	neg_2
24	pos_375
25	neg_375
26	pos_5
27	neg_5
28	pos_625
29	neg_625
30	pos_75
31	neg_75
32	pos_875
33	neg_875
34	pos_1
35	neg_1

	Features VADER
1	Number_of_word
2	Number_of_sentences
3	Number_of_caps
4	Number_of_punctuation
5	Number_of_nouns
6	Number_of_verbs
7	Number_of_adjectives
8	Number_of_adverbs
9	Average_word_length
10	Average_sentence_length
11	Redundancy_score
12	Readability_score
13	Lexical_diversity
14	Sentiment_score
15	positive_score
16	negative_score
17	neutral_score
18	positive_words
19	negative_words
20	neutral_words
21	polarity_shifters
22	intensity_modifiers
23	negations
24	emoticons

	Features Combined
1	Number of words
2	Number of sentences
3	Number of caps
4	Number of punctuation
5	Number of nouns
6	Number of verbs
7	Number of adverbs
8	Number of adjectives
9	Average word length
10	Average sentence length
11	Redundancy score
12	Readability score
13	Lexical diversity
14	positive score
15	Review ambiguity
16	Number of objective words
17	polarity shifters
18	neutral score
19	negations
20	Sentiment score
21	Review intensity
22	Number of positive words
23	Number of negative words
24	intensity modifiers
25	emoticons
26	neg 125
27	pos 125
28	pos 25
29	neg 25
30	pos 375
31	neg 375
32	neg 5
33	pos 5
34	pos 625
35	neg 625
36	neg 75
37	pos 75
38	pos 875
39	neg 875
40	pos 1
41	neg 1

Appendix 2:

Table 6: T-Test Results for Comparison of Fake and Real Reviews (VADER Lexicon)

Feature	T-Statistic	P-Value
Number of words	-17,282	1,84851E-66
Number of sentences	-18,685	2,75474E-77
Number of caps	-10,209	2,06043E-24
Number of punctuation	-18,929	2,99644E-79
Number of nouns	-18,917	3,7634E-79
Number of verbs	-15,407	2,89626E-53
Number of adjectives	-14,887	7,24274E-50
Number of adverbs	-12,083	1,68101E-33
Average word length	-3,311	0,000932547
Average sentence length	-2,265	0,023520351
Redundancy score	-4,686	2,8018E-06
Readability score	-4,825	1,41048E-06
Lexical diversity	6,906	5,11105E-12
Sentiment score	0,753	0,451536566
Positive score	11,573	7,02192E-31
Negative score	0,394	0,693625851
Neutral score	-13,068	7,15567E-39
Positive words	-6,426	1,3368E-10
Negative words	-10,645	2,13849E-26
Neutral words	-17,913	3,17503E-71
Polarity shifters	-19,568	1,63752E-84
Intensity modifiers	-4,972	6,69565E-07
Negations	-14,476	2,88655E-47
Emoticons	-2,781	0,005418027

Table 7: T-Test Results for Comparison of Fake and Real Reviews (SentiWordNet Lexicon)

Feature	T-Statistic	P-Value
Number of words	-17,282	1,84851E-66
Number of sentences	-18,685	2,75474E-77
Number of caps	-10,209	2,06043E-24
Number of punctuation	-18,929	2,99644E-79
Number of nouns	-18,917	3,7634E-79
Number of verbs	-15,407	2,89626E-53
Number of adjectives	-14,887	7,24274E-50
Number of adverbs	-12,083	1,68101E-33
Average word length	-3,311	0,000932547
Average sentence length	-2,265	0,023520351
Redundancy score	-4,686	2,8018E-06
Readability score	-4,825	1,41048E-06
Lexical diversity	6,906	5,11105E-12
Sentiment score	12,462	1,60554E-35
Review intensity	-11,458	2,64685E-30
Review ambiguity	-17,076	6,17195E-65
Number of positive words	-9,219	3,28897E-20
Number of negative words	-18,447	2,17854E-75
Number of objective words	-17,854	9,03464E-71
Pos 125	-13,987	2,96295E-44
Neg 125	-14,395	9,28049E-47
Pos 25	-10,031	1,26339E-23
Neg 25	-12,912	5,34576E-38
Pos 375	-7,380	1,64231E-13
Neg 375	-13,147	2,54183E-39
Pos 5	-6,756	1,45336E-11
Neg 5	-16,940	6,0914E-64
Pos 625	-1,954	0,050671311
Neg 625	-15,521	4,96318E-54
Pos 75	-1,259	0,208151238
Neg 75	-7,054	1,79259E-12
Pos 875	1,901	0,057376815
Neg 875	-2,568	0,010235479
Pos 1	-	-
Neg 1	-1,838	0,066113994

Appendix 3:

	SVM Average SentiWordNet	NB Average SentiWordNet	SVM Average VADER	NB Average VADER	SVM Average VADER / Senti	NB Average VADER / Senti
41 Ft					57,42	42,95
40 Ft					57,42	42,95
39 Ft					57,39	42,91
38 Ft					57,47	42,78
37 Ft					57,71	42,44
36 Ft					57,72	42,44
35 Ft	56,92	42,54			57,75	42,47
34 Ft	56,92	42,54			57,56	42,47
33 Ft	52,70	42,57			57,74	42,28
32 Ft	52,73	42,29			57,48	41,98
31 Ft	52,70	42,29			56,87	41,87
30 Ft	49,15	42,25			56,80	41,76
29 Ft	49,11	42,33			56,68	41,75
28 Ft	49,08	42,30			56,91	41,69
27 Ft	49,17	42,07			56,79	41,75
26 Ft	49,04	42,31			56,96	41,64
25 Ft	48,11	41,38			56,80	41,69
24 Ft	48,47	41,50	55,17	42,16	56,74	41,97
23 Ft	48,33	41,43	52,71	41,84	56,23	41,95
22 Ft	48,07	41,46	52,94	41,78	56,31	41,72
21 Ft	47,55	41,42	51,95	41,70	56,41	41,46
20 Ft	47,67	41,46	51,93	41,84	56,16	41,40
19 Ft	47,34	41,55	51,94	41,73	56,15	41,30
18 Ft	47,43	41,35	51,90	41,72	56,08	41,63
17 Ft	47,32	41,35	51,89	41,75	56,04	41,59
16 Ft	47,32	41,23	51,47	41,46	56,00	41,39
15 Ft	47,10	41,50	51,25	41,54	55,81	41,56
14 Ft	46,49	41,31	51,25	41,25	55,76	41,80
13 Ft	46,33	41,44	51,24	41,31	55,73	42,19
12 Ft	46,00	41,42	51,92	42,41	55,67	42,04
11 Ft	45,87	41,48	52,01	42,20	55,11	42,88
10 Ft	45,38	41,43	49,23	42,37	54,49	42,15
9 Ft	44,46	41,47	49,01	41,85	54,19	41,78
8 Ft	44,49	41,30	45,28	41,40	54,55	41,69
7 Ft	43,67	41,34	45,67	41,25	53,60	41,19
6 Ft	44,02	41,24	45,43	41,13	53,24	41,22
5 Ft	42,86	40,62	45,25	41,12	53,24	41,40
4 Ft	41,99	40,51	45,11	40,34	53,25	41,14
3 Ft	41,90	40,37	43,50	39,12	53,41	40,48
2 Ft	40,96	39,42	43,20	48,39	56,66	55,42
1 Ft	37,95	46,28	41,55	46,28	48,80	46,28
Maximum	**56,915**	**46,2825**	**55,165**	**48,39**	**57,75**	**55,42**

Appendix 4:

	SVM Accuracy SentiWordNet	NB Accuracy SentiWordNet	SVM Accuracy VADER	NB Accuracy VADER	SVM Accuracy VADER / Senti	NB Accuracy VADER / Senti
41 Ft					59,12	55,38
40 Ft					59,12	55,38
39 Ft					59,10	55,36
38 Ft					59,12	55,31
37 Ft					59,40	55,14
36 Ft					59,40	55,14
35 Ft	59,11	55,24			59,38	55,17
34 Ft	59,11	55,24			59,38	55,17
33 Ft	58,62	55,26			59,55	55,10
32 Ft	58,64	55,12			59,33	54,95
31 Ft	58,62	55,12			58,52	54,90
30 Ft	57,74	55,1			58,79	54,86
29 Ft	57,71	55,14			58,57	54,86
28 Ft	57,69	55,14			58,67	54,81
27 Ft	57,71	55,02			58,50	54,86
26 Ft	57,62	55,14			58,74	54,81
25 Ft	57,1	54,69			58,64	54,83
24 Ft	57,36	54,76	57,4	54,97	58,60	54,98
23 Ft	57,24	54,74	56,71	54,9	58,14	54,98
22 Ft	57,05	54,76	56,88	54,86	58,19	54,86
21 Ft	56,98	54,74	56,52	54,83	57,74	54,74
20 Ft	57,05	54,76	56,52	54,9	57,64	54,74
19 Ft	56,98	54,81	56,55	54,88	57,57	54,67
18 Ft	57,05	54,71	56,5	54,86	57,60	54,83
17 Ft	56,98	54,71	56,55	54,88	57,64	54,81
16 Ft	56,98	54,67	56,29	54,71	57,57	54,71
15 Ft	56,86	54,79	55,86	54,74	57,21	54,79
14 Ft	56,88	54,69	55,9	54,62	57,14	54,90
13 Ft	56,81	54,76	55,81	54,62	57,21	55,07
12 Ft	56,67	54,74	56,02	55,17	57,17	55,00
11 Ft	56,6	54,76	56,05	55,07	56,76	55,38
10 Ft	56,4	54,74	55,74	55,05	56,69	55,02
9 Ft	56	54,71	56,16	54,86	56,43	54,83
8 Ft	56,05	54,62	55,88	54,67	56,57	54,81
7 Ft	55,64	54,67	56,12	54,6	54,38	54,60
6 Ft	55,86	54,64	56,02	54,57	54,33	54,62
5 Ft	55,33	54,35	56,14	54,57	54,33	54,71
4 Ft	54,88	54,31	56,12	54,21	54,60	54,57
3 Ft	54,9	54,24	55,31	53,69	54,19	54,24
2 Ft	54,52	53,83	55,19	51,45	52,64	52,76
1 Ft	53,31	52,38	54,6	52,38	51,60	52,38
Maximum	**59,11**	**55,26**	**57,4**	**55,17**	**59,55**	**55,38**

Appendix 5:

	SVM Precision SentiWordNet	NB Precision SentiWordNet	SVM Precision VADER	NB Precision VADER	SVM Precision VADER / Senti	NB Precision VADER / Senti
41 Ft					59,76	68,94
40 Ft					59,76	68,94
39 Ft					59,73	68,82
38 Ft					59,73	68,98
37 Ft					60,08	68,58
36 Ft					60,07	68,58
35 Ft	60,12	69,75			60,01	68,84
34 Ft	60,12	69,75			60,15	68,84
33 Ft	63,11	69,96			60,33	69,02
32 Ft	63,14	69,53			60,12	68,67
31 Ft	63,11	69,46			59,05	68,69
30 Ft	65,82	69,4			59,57	68,64
29 Ft	65,78	69,67			59,26	68,71
28 Ft	65,71	69,98			59,28	68,23
27 Ft	65,62	69,52			59,06	68,71
26 Ft	65,42	69,82			59,37	68,66
25 Ft	64,75	68,42			59,30	68,73
24 Ft	65,34	68,99	58,11	67,77	59,26	69,40
23 Ft	65,04	69,25	58,29	68,91	58,78	69,64
22 Ft	64,59	69,47	58,48	68,49	58,81	69,09
21 Ft	65,41	69,33	58,39	68,65	57,98	68,85
20 Ft	65,52	69,47	58,4	68,99	57,96	69,49
19 Ft	65,98	69,85	58,45	69,55	57,85	68,66
18 Ft	66,17	69,51	58,38	69,1	57,92	69,42
17 Ft	66,01	69,51	58,49	69,39	58,03	69,20
16 Ft	66,01	69,63	58,21	68,26	57,93	68,94
15 Ft	65,85	69,79	57,46	68,04	57,45	68,90
14 Ft	68,12	69,28	57,55	68,08	57,36	69,37
13 Ft	68,14	69,72	57,36	67,58	57,48	69,41
12 Ft	68,18	69,24	57,4	69,42	57,44	69,16
11 Ft	68,04	69,23	57,4	69,26	57,06	69,57
10 Ft	68,35	69,16	58,53	67,5	57,24	68,70
9 Ft	68,37	68,19	59,89	67,86	56,96	68,01
8 Ft	68,82	67,65	64,35	67,78	57,02	68,23
7 Ft	68,2	68,28	64,9	67,59	54,18	68,09
6 Ft	69,1	68,75	64,91	68,1	54,21	68,38
5 Ft	68,83	67,87	66,35	68,17	54,21	68,79
4 Ft	67,34	67,89	66,67	67,23	54,56	67,95
3 Ft	68,4	67,45	65,29	65,91	53,98	66,60
2 Ft	68,66	66,59	65,31	51,3	51,85	52,08
1 Ft	68,56	52,95	65,78	52,95	51,44	52,95
Maximum	**69,1**	**69,98**	**66,67**	**69,55**	**60,33**	**69,64**

Appendix 6:

	SVM Recall SentiWordNet	NB Recall SentiWordNet	SVM Recall VADER	NB Recall VADER	SVM Recall VADER / Senti	NB Recall VADER / Senti
41 Ft					54,05	18,42
40 Ft					54,05	18,42
39 Ft					54,00	18,42
38 Ft					54,20	18,13
37 Ft					54,29	17,79
36 Ft					54,34	17,79
35 Ft	52,42	17,36			54,48	17,70
34 Ft	52,42	17,37			53,86	17,70
33 Ft	40,05	17,31			54,05	17,31
32 Ft	40,1	17,07			53,72	17,03
31 Ft	40,05	17,12			53,67	16,83
30 Ft	30,94	17,07			52,85	16,69
29 Ft	30,89	17,07			52,95	16,64
28 Ft	30,89	16,88			53,48	16,79
27 Ft	31,13	16,74			53,48	16,64
26 Ft	31,03	16,98			53,48	16,50
25 Ft	29,78	16,21			53,19	16,55
24 Ft	30,02	16,12	50,89	17,75	53,09	16,64
23 Ft	29,98	15,88	45,04	16,69	52,52	16,50
22 Ft	29,83	15,83	45,32	16,79	52,66	16,40
21 Ft	28,3	15,83	43,21	16,59	54,00	16,12
20 Ft	28,44	15,83	43,17	16,64	53,43	15,73
19 Ft	27,53	15,78	43,12	16,21	53,57	15,97
18 Ft	27,58	15,64	43,12	16,4	53,29	16,12
17 Ft	27,48	15,64	42,97	16,31	53,05	16,16
16 Ft	27,48	15,4	42,35	16,4	53,09	15,97
15 Ft	27,19	15,73	42,69	16,64	53,29	16,26
14 Ft	24,7	15,68	42,59	16,16	53,29	16,40
13 Ft	24,41	15,68	42,78	16,5	53,05	16,98
12 Ft	23,84	15,88	44,27	17,31	52,95	16,88
11 Ft	23,69	15,97	44,46	17,07	52,13	17,99
10 Ft	22,69	15,92	37,17	18,23	50,41	17,27
9 Ft	21,15	16,45	35,44	17,22	50,07	17,03
8 Ft	20,96	16,45	24,94	16,55	50,84	16,79
7 Ft	19,95	16,21	25,28	16,4	52,52	16,07
6 Ft	20,05	15,83	24,84	15,97	51,56	15,97
5 Ft	18,32	15,3	23,65	15,92	51,56	16,07
4 Ft	17,7	15,11	23,21	15,16	51,08	16,07
3 Ft	17,03	15,11	21,29	13,91	52,33	15,68
2 Ft	15,44	14,05	20,77	43,65	64,60	60,77
1 Ft	10,98	36,55	17,79	36,54	44,46	36,55
Maximum	**52,42**	**36,55**	**50,89**	**43,65**	**64,60**	**60,77**

Appendix 7:

	SVM F1-Score SentiWordNet	NB F1-Score SentiWordNet	SVM F1-Score VADER	NB F1-Score VADER	SVM F1-Score VADER / Senti	NB F1-Score VADER / Senti
41 Ft					56,76	29,07
40 Ft					56,76	29,07
39 Ft					56,73	29,06
38 Ft					56,83	28,71
37 Ft					57,04	28,26
36 Ft					57,06	28,26
35 Ft	56,01	27,8			57,11	28,16
34 Ft	56,01	27,8			56,83	28,16
33 Ft	49	27,76			57,02	27,68
32 Ft	49,05	27,42			56,74	27,29
31 Ft	49	27,47			56,23	27,04
30 Ft	42,09	27,41			56,01	26,85
29 Ft	42,04	27,43			55,93	26,80
28 Ft	42,02	27,2			56,23	26,94
27 Ft	42,23	26,98			56,13	26,80
26 Ft	42,1	27,31			56,27	26,60
25 Ft	40,8	26,21			56,08	26,67
24 Ft	41,14	26,13	54,26	28,13	56,01	26,85
23 Ft	41,04	25,83	50,81	26,87	55,47	26,68
22 Ft	40,81	25,78	51,07	26,96	55,57	26,51
21 Ft	39,5	25,77	49,67	26,73	55,92	26,12
20 Ft	39,67	25,78	49,64	26,81	55,60	25,66
19 Ft	38,85	25,74	49,63	26,29	55,63	25,91
18 Ft	38,93	25,53	49,6	26,51	55,51	26,16
17 Ft	38,8	25,53	49,54	26,41	55,42	26,21
16 Ft	38,81	25,22	49,03	26,45	55,41	25,93
15 Ft	38,49	25,68	48,98	26,74	55,29	26,31
14 Ft	36,25	25,58	48,95	26,12	55,25	26,53
13 Ft	35,95	25,61	49,01	26,52	55,18	27,28
12 Ft	35,32	25,83	49,99	27,72	55,10	27,14
11 Ft	35,15	25,95	50,11	27,4	54,49	28,58
10 Ft	34,07	25,89	45,47	28,7	53,61	27,60
9 Ft	32,31	26,51	44,53	27,47	53,29	27,23
8 Ft	32,13	26,47	35,95	26,6	53,75	26,94
7 Ft	30,87	26,2	36,38	26,4	53,34	26,00
6 Ft	31,08	25,73	35,93	25,87	52,85	25,89
5 Ft	28,94	24,97	34,87	25,82	52,85	26,05
4 Ft	28,03	24,72	34,44	24,74	52,76	25,99
3 Ft	27,27	24,69	32,12	22,97	53,14	25,39
2 Ft	25,21	23,21	31,51	47,16	57,53	56,09
1 Ft	18,93	43,25	28,01	43,24	47,70	43,25
Maximum	**56,01**	**43,25**	**54,26**	**47,16**	**57,53**	**56,09**

Appendix 8:

Rank	Feature
35	pos 875
34	neg 375
33	negations
32	neg 625
31	neg 5
30	pos 375
29	pos 625
28	neg 25
27	pos 25
26	neg 125
25	pos 5
24	pos 125
23	Number of negative words
22	Number of positive words
21	Number of adjectives
20	Number of adverbs
19	intensity modifiers
18	Number of verbs
17	polarity shifters
16	Number of objective words
15	Review ambiguity
14	Number of sentences
13	Number of caps
12	Number of nouns
11	Average word length
10	Lexical diversity
9	Average sentence length
8	Review intensity
7	Sentiment score
6	Readability score
5	neutral score
4	Number of words
3	Redundancy score
2	positive score
1	Number of punctuation

Appendix 9:

Rank	Feature
33	negations
32	neg 625
31	neg 5
30	pos 375
29	pos 625
28	neg 25
27	pos 25
26	neg 125
25	pos 5
24	pos 125
23	Number of negative words
22	Number of positive words
21	Number of adjectives
20	Number of adverbs
19	intensity modifiers
18	Number of verbs
17	polarity shifters
16	Number of objective words
15	Review ambiguity
14	Number of sentences
13	Number of caps
12	Number of nouns
11	Average word length
10	Lexical diversity
9	Average sentence length
8	Review intensity
7	Sentiment score
6	Readability score
5	neutral score
4	Number of words
3	Redundancy score
2	positive score
1	Number of punctuation