Fraud Detection in E-commerce Using Natural Language Processing

by

Iftekhar Kabir
18201106
Marium Khan Momo
18301069
Tahsin Tazrian
19101520

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

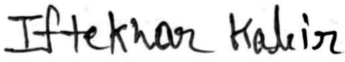
Department of Computer Science and Engineering
Brac University
January 2023

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



Iftekhar Kabir
18201106

Marium Khan Momo
18301069

Tahsin Tazrian
19101520

Approval

The thesis/project titled "Fraud Detection in E-commerce Using Natural Language Processing" submitted by

1. Iftekhar Kabir (18201106)
2. Marium Khan Momo(18301069)
3. Tahsin Tazrian (19101520)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 17, 2023.

Examining Committee:

Supervisor:
(Member)

Farig Yousuf Sadeque

Dr. Farig Yousuf Sadeque
Assistant Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Robiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi
Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Electronic commerce sometimes referred to as e-commerce is a type of business that enables both businesses and private individuals to purchase and sell products and services online. E-commerce in Bangladesh is thriving from the last decade, especially during the coronavirus pandemic with the growth of online sales. Digital commerce is currently struggling to regain trust after allegations of annexation and fraud surfaced against a few firms in recent months. Over 11.48% clients of the internet business area were beguiled last year from different web based business and Facebook trade (business) sites. Fake reviews are one of the most prominent fraudulent activities in this field. When we try to buy anything online or book any hotel from an app or a ride from any ride sharing app we heavily rely on the reviews of past customers.It makes the decision making process easier. This is why, with the ongoing development of e-commerce platforms online reviews are seen as essential to upholding a company's reputation. Generally a positive feedback from a customer gathers the attraction of many searching for the same product. For this reason, many e-commerce sites are generating fake reviews to attract more customers towards them. Detecting fake reviews is an ongoing research area. As all the reviews are not trustworthy and honest, it is crucial for us to develop techniques for detecting fake reviews. We are proposing a machine learning approach to generate and detect fake reviews.We used Natural Language Processing(NLP) to extract meaningful features from a text for detecting fraud reviews. Therefore, in this study, we present a comprehensive and effective framework that enhances the efficacy of fake review identification using Support Vector Machine(SVM) and Logistic Regression machine learning algorithm among several machine learning algorithms for detecting fake reviews.

Keywords: Fake Review; Machine Learning; Support Vector Machine; Logistic Regression; Detection; BiDirectional Long Short Term Memory

Acknowledgement

BRAC University provided support to facilitate the completion of this study. First and foremost, we thank the Great Allah for sparing us damage during the COVID-19 pandemic and enabling us to finish our research on schedule. We would want to take this opportunity to thank our supervisor, Dr. Farig Yousuf Sadeque sir for everything that he has done to help us and for letting us work for him. And last, many thanks to our parents for their thoughtful prayers and help.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

CG Computer Generated

CNN Convolutional Neural Network

LR Logistic Regression

$LSTM$ Long Short-Term Memory

NB Naive Bayes

NLP Natural Language Processing

OR Original Review

RNN Recurrent Neural Network

SSL Semi-supervised learning

SVM Support Vector Machine

$TF - IDF$ Term Frequency-Inverse Document Frequency

Chapter 1

Introduction

E-commerce is becoming the most popular platform for people all over the world because of its easy to access and user-friendly interface. E-commerce sites are being dominated by the "phenomenon of fake". Reviews may have an impact on customers in a variety of industries, but they are especially important in the world of online commerce where customers use comments and reviews to determine whether or not to buy products and services. Generally service providers consult with their clients to provide their valuable feedback on the items they have purchased from them so that it can improve their reputation and make their business trustworthy to other customers. Platforms for social networking and online shopping have assimilated into today's society. As a result, the amount of data on the Web is rapidly increasing, but further research is needed to determine its quality. People now use social media platforms to express themselves and remain in touch by talking about the news, weighing in on political issues, viewing movies or buying things[21]. Unfortunately, this online data can spread rapidly and easily and it is not difficult to tamper with it. Today particularly during emergency situations like the COVID-19 epidemic that we currently experienced, fake news and misleading information have grown more and more obvious. Generally, in Fraud detection it wants to get to a place where it build like a cyclical or feedback loop of sorts that gets a better fraud detection as time passes. So what really want focus on, in fraud detection, especially when it has to do with content or things that are readable by an NLP, which it want to start with the idea of understanding , so that ends up being a text mining or NLP largely natural language understanding. So, we want to have an understanding of what the specific piece of content is saying and start understanding how it might relate to fraud. Numerous research have shown that in order to make the COVID-19's curvature flatter, the curve of rumors and false information about the virus needs to be flattened. A significant example of where phony reviews and profiles are prevalent is on e-commerce platforms, where there is an increasing number of new products and fierce competition amongst businesses. Online reviews can be fake, yet they are nonetheless important for both customers and businesses. Consumers must inspect them in order to judge the products' quality and whether or not to purchase them. On the other hand, businesses require internet reviews to enhance their offerings and increase sales. Fact-checking systems that can recognize bogus reviews or fake users must be created in order to improve the transparency and quality of the data otherwise, the Online world would be a sea of unusually pervasive rumors and false information. Clearly, this task is not a simple one. Despite significant

work by scholars in this field, there is still room for improvement. We must first acknowledge the properties of such data and the fraudulent users' behavior patterns in order to fully comprehend the issue. Positive internet evaluations, according to 90% of consumers who remembered reading them, had an influence on their purchasing decisions. In addition, 86% of them claimed that unfavorable online reviews influenced their purchasing choices[33]. Therefore, it is imperative to have a reasoning system that can identify these reviews especially when taking into account the following factors: 1.Customers who read fake reviews are led to believe that they are completely informed about the goods they are purchasing[33]. 2. To prevent customers from giving a product a positive review, some phony reviews are written[33]. 3. Many false reviews are published to either enhance or degrade the product[33]. 4. Because of the paucity of training data that takes into account the reviewers' behavioural patterns, pure supervised learning algorithms may not be the best option. The suggested method makes use of supervised learning to analyze the review's sentiment, which is taken into account by one KB rule[33] using NLP which is natural language processing to try to predict fake reviews. Now, this is obviously something that has come to the fore more recently than in the past. But, we venture to speculate that fake reviews has been something in and around the globe pretty much forever there's that famous quote that a lie can make it around the world. Before the truth has time to put its pants on so that has always been the case and it will probably always be the the case so the real question is this can we utilize new technology data science modeling different things like that to try to predict if the review is genuine or not whether it is a hoax, etc. before or once social media starts to spread it. The detection of fake reviews in natural language processing is a crucial but difficult problem (NLP)[32]. In addition to greatly increasing information accessibility, the quick development of social networking sites has also sped up the dissemination of false evaluations. Because of this, the impact of bogus reviews has been expanding, sometimes even affecting the offline world and endangering public safety. Given the vast amount of online material, automatic fake review identification is a realistic NLP challenge that will benefit all online content producers by minimizing the amount of time and effort needed for people to recognize and stop the propagation of phony reviews. In this study, we discuss the difficulties in detecting false reviews as well as related issues[17]. We comprehensively examine, contrast and evaluate the potentials and restrictions of the task formulations, datasets and NLP alternatives that have been created for this task. We describe prospective research directions based on our findings, including more accurate, thorough, fair and useful detection methods. We also emphasize the distinction between the detection of false reviews and other similar tasks, as well as the significance of NLP tools for fake review detection[17].

1.1 Research Problem

Computer science's natural language processing (NLP) discipline examines how people and machines communicate. Alan Turing devised an intelligence test in a 1950s article that is now known as the Turing test. Results have been achieved in the areas of language modeling, parsing and natural language challenges using more contemporary methodologies, such as deep learning[24]. In order to detect false product reviews at one of the top virtual merchants in the world, we tested state-of-

the-art algorithms for natural language processing as part of our research project. All of our models were developed using the amazon reviews dataset, which provides explicit labels identifying false reviews[29]. Deceptive opinions, spam reviews and spam opinions are frequently used as definitions for fake reviews. Their authors may also be referred to as spammers. The three categories of spam opinions often known as phony reviews are as follows: 1. Users that publish reviews with untrue comments do so in an effort to harm the reputation of a company or product or to support it. These evaluations are referred to as phony or deceptive reviews and since legitimate and fraudulent reviews are similar to one another, it can be difficult to identify them by reading alone. 2. Only those who comment on the brand of the products are described in reviews of that brand. 3. Non-reviews that are pointless don't provide a real viewpoint or are merely ads. The last two varieties referred to as disruptive spam opinions pose little risk and are easily distinguishable by anyone who reads them. We must take into account the following two samples of reviews from a Yelp Chi real-life public dataset in order to describe and comprehend the nature of phony reviews. While the second review is bogus, the first is real. Review 1: "I like staying here. The staff is really kind and makes you feel right at home. Fantastic location and excellent lodging." Review 2: "What an incredible hotel. The employees are excellent and thoughtful. Benefits like the free bike rentals are fantastic. The building's refurbishment and history are both quite fascinating. I appreciate you making my stay so wonderful with your assistance[34]."

1.2 Related Works

User reviews have a significant influence on how much money a company makes in e-commerce. Before choosing any goods or service, online users rely on reviews. As a result, the legitimacy of online evaluations is essential for businesses and has a direct impact on their profitability and reputation[19]. Because of this, some companies pay spammers to publish phony reviews. These fraudulent reviews take advantage of consumer purchasing choices[19]. As a result, during the past twelve years a lot of research has been done on how to spot false reviews. However, a survey that can evaluate and summarize the current methods is still lacking. This survey study summarizes the existing datasets and their gathering techniques in order to address the issue and describes the task of detecting fraud review. It examines the feature extraction methods currently in use. Additionally, it provides a comprehensive summary of the available methodologies and assesses them in two groups traditional statistical machine learning techniques and deep learning approaches to find any shortcomings. Additionally, we carry out a benchmark research to assess the effectiveness of various transformers and neural network models that have not previously been used to the identification of fraudulent reviews. According to research findings on two benchmark datasets, RoBERTa surpasses state-of-the-art algorithms in a mixed domain with the maximum accuracy of 91.2% for the deception dataset and this performance may be used as a benchmark for more study[31]. We conclude by highlighting the research area's present deficiencies and potential future directions. In regards to product reviews, Jindal and Liu initially raised the issue of bogus reviews identification in 2008. Multiple ways have been used to address the issue, including classifying reviews as fake or genuine by examining the review content, user behavioral traits or sentiment analysis techniques. Many researchers have

worked to develop cutting-edge methods for determining whether reviews are real or spam. As stated in reference, several detection levels should be taken into account; the primary ones are as follows: review content and user behavior-based detection, review variances among rating-based detection and review content-based detection. Supervised models use the first two categories, however it may be challenging to achieve correct results because training the models may need a huge dataset[35]. To improve the overall performance, the authors suggested taking into account the third category. In-depth research has recently been conducted on the development of machine learning-based spam detection systems. Reference presents two distinct methods for identifying false reviews, where new semantic characteristics were extracted and classified using supervised learning methods using a real Yelp review dataset. In contrast to n-grams, the research suggests a collection of behavioral traits and demonstrates how include them improves accuracy. The length of the reviews, the intervals between reviews, and the ratio of each good to negative review were all behavioral factors that the authors said helped them achieve high accuracy. For classification, Support Vector Machine (SVM) has been employed. A feature framework for fake review identification has been developed using a classification algorithm. The study resolves the issue by pulling data from the actual Yelp dataset in order to produce a brand-new dataset for the consumer electronics area. The framework does employ two categories of features for feature extraction: review-centric features which are concerned with the review as a text and user-centric features which are focused on the reviewers' behavioral patterns, such as their personal, social and review-related activities. As well as verbal characteristics like the length or content of the review, nonverbal behavioral traits like the number of likes or dislikes, the average posting frequency and review updates among others can be useful in identifying fraudulent reviews. The study proposed non-machine learning methods that look at relationships between reviewers, reviews and content similarity, such as graph-based or pattern matching methods. Additionally, the research in Reference shows that incorporating the reviewers' behavioral attributes increased accuracy by about 20% when compared to n-grams, but they also noted that additional information such as IP addresses, user logs or session lengths might be employed to enhance the outcomes. Other strategies suggested a phony review recognition system by capturing questionable review time intervals. Using private information such as IP addresses and MAC addresses can improve the performance of the false detection system. The issue that there are no dataset quality requirements that can be used to attain 100% accuracy in determining whether a review is false or not is shared by both articles. Additionally, it exhibits a variety of qualities that make it easier to identify phony reviews. These elements, which include the reviewers, their reviews and the capability of repeating reviews are mostly used to identify similarities in review content. Additionally, the things that have already been examined as well as how frequently they are done are taken into account. The direct effect of internet evaluations on the box office earnings of a particular film, indicating that consumers occasionally use these online services to determine whether or not to watch a film. One of the main problems when using machine learning techniques is the imbalance between the states of the two classes, false and real information. Getting labels for misleading information might be challenging. These are frequently gathered manually by professionals, educated volunteers or employees of Amazon Mechanical Turk[14]. The procedure entails a lot of manual work and the assessors might not

be able to categorize every piece of false information they encounter. Rule-based procedures, in contrast are acknowledged as a white box providing traceability and transparency for important judgments that call for a deeper level of explanation than is typically provided by machine learning approaches. Additionally, the choice between a rule-based system and a system using machine learning rely on the nature of the issue at hand and mostly involves balancing the trade-offs between efficiency, training costs and comprehension.

Chapter 2

Literature Review

2.1 Natural Language Processing(NLP)

The term "Natural Language Processing," which stands for artificial intelligence, refers to a computer system's capacity to comprehend spoken language. Computer science's field of "Natural Language Processing" is concerned with human-computer interaction. Users of NLP may ask inquiries about any topic and receive a quick answer in a matter of seconds. This system responds to the query with natural language responses. The system provides precise responses to the queries without providing extra or undesired information. This method scales other language-related activities as well as enables computers to converse with people in their native tongue. Speech recognition, language comprehension and language production are typically difficult aspects of natural language processing. NLG deals with generating spoken or written language from unstructured data. This is perhaps the best known database to be found in the pattern recognition literature. NLU deals with Understanding the input given by the user as a part of natural language. Basically, it is exploring a dataset of how many rows are there in the database, how many labels and how many data are adding and missing. Applying automatic detection NLP approaches emphasize lexical features (i.e., textual properties) such keywords or -s, punctuation, n-grams, latent themes and semantic similarity when processing reviews as textual data[1] . Along with NLP we can also consider human detectors to assist us in detecting fake reviews . Harris[2] suggested a hybrid method in which the output of two machine learning classifiers was combined with psycho-linguistic features that had been retrieved algorithmically.This hybrid technique improved machine performance by 0.2 percentage points, demonstrating that human input may lead to a modest (but statistically significant) advantage over a wholly automated strategy. Humans might concur or disagree with the computer decision.

2.2 Synthetic Review Detection

Due to the scarcity of review spam samples and the difficulties in categorizing them, the majority of the datasets used in the earlier research are synthetically constructed [3]. Synthetic reviews that use sentence transplants have a subtler semantic incoherence between sentences than natural writings. Because these synthetic datasets are not always reflective of real-world review spam, developing and testing classifiers based on them can be challenging. For instance, the derived characteristics

and outcomes while evaluating the artificial AMT dataset used in [4,5,6] and Yelp's filtered reviews dataset substantially varied, particularly in the case of n-gram text features[7]. When these datasets' classification performance is compared, it is clear that using the Yelp review dataset as a benchmark, the classifier got an accuracy of 65%, but while using synthetic reviews it achieved 87% accuracy. This 22% decline in accuracy suggests that synthetic reviews vary from real-life fake reviews in terms of differentiating characteristics and that AMT does not correctly represent real-world spam reviews in its reviews.

2.3 Supervised Learning Method for Detecting Fake Review

A combination of machine learning and artificial intelligence supervised learning is sometimes referred to as supervised machine learning. It shines out because it trains algorithms that properly classify data or forecast occurrences using labeled datasets.During the cross validation process, weights are modified as input data are put into the model until the model is well fitted.With the use of supervised learning, businesses may find sustainable solutions to a wide range of real-world issues. In[21] Their proposed model is divided into four stages. The first step which involves data collecting and data preparation uses both labeled and unlabeled datasets and preprocesses them.Preprocessing is done using both labeled and unlabeled data in Natural Language Processing (NLP) methods such as stop word and punctuation removal, changing to lowercase letters in english and stemming. In the second stage the Active Learning Algorithm is used to gradually label all of the unlabeled data while the learner evaluates the dataset's accuracy by comparing the probability difference with a threshold value for accurate classification. The third step covers the technique for selecting features which includes the n-grams, TF- IDF and Word Embedding techniques. They have used both TF-IDF and n-grams approaches for ordinary machine learning and TF-IDF (Term-Frequency, Inverse Term Frequency), Word Embeddings (Word2Vec) and LSTM techniques for deep learning to display text as a set of numbers. In the last stage of their proposed methodology reviews are labeled as spam and ham using both traditional machine learning and deep learning classifiers. SVM, KNN and Naive Bayes (NB) classifiers are used for detecting fraud reviews instead of frequently used machine learning techniques. They have used Multilayer Perceptron (MLP), CNN and RNN deep learning approaches for identifying fraud (they have utilized LSTM, which is a version of RNN).

2.4 Unsupervised Learning Method for Detecting Fake Review

The use of supervised learning isn't always appropriate due to the challenge of creating accurately labeled datasets of fake review . This issue is solved by using unsupervised learning methods as they don't require labeled data. For the purpose of identifying untruthful reviews, a novel unsupervised text mining model was created, incorporated it into a semantic language model and compared it to supervised learning techniques[8]. Their research provides an approximation method

for determining the level of authenticity for reviews based on the duplicate identification findings estimating the similarity of semantic contents among reviews that used a Semantic Language Model.They created a high-order concept of association mining in addition to performing unsupervised fake review detection to extrapolate idea association knowledge that is context-sensitive. The experimental findings demonstrate the effectiveness of semantic language modeling and text mining-based computational models for the identification of fraud reviews, as well as the high proportion of duplicate fake reviews being found that may be attained using unsupervised approaches.

2.5 Text Reuse Detection

Text reuse is the practice of repeatedly using text from earlier works. Text reuse detection has been the subject of in-depth research investigations in the context of online searches.For example the identification of duplicate or near-duplicate documents [9,10]. The resemblance[9] determines whether two (web) documents are almost identical, that is, they differ only in their alterations of the same information as in - minor corrections, web-master signature, formatting, logo, capitalization etc. When the likeness is close to 1, it is likely that the two papers are identical to one another. The resemblance is a value within 0 and 1 which is specifically detailed below. They have presented a technique that can exclude almost similar documents from a library of hundreds of millions of files by computing individually for each document a vector of characteristics just under 50 bytes long and examining just these vectors instead of complete documents.

2.6 Feature Engineering for fake review detection

Features are either developed or retrieved from data in feature engineering. Various elements that may be extracted from reviews have been used in previous research with words from the review's text being the most common. The bag of words method often employed for this in which each review's characteristics are either a single word or a small set of terms that may be found inside the review's content. Less frequently, syntactical and lexical characteristics[11] or features describing the activities of reviewers have been included in research as extra components of the items, reviewers and reviews.In general, training a classifier with many types of features has led to greater performance than training it with a single type of feature. According to a research by Mukherjee et al. [12], using the unusual behavioral traits of reviewers outperformed using the reviews' language features.

Chapter 3

Background Studies

Nowadays, it seems like there are fake reviews everywhere, making it difficult for shoppers to determine whether goods or companies are reliable. There is always a chance that the reviews you are reading are fraudulent, whether you are shopping on Amazon, researching a restaurant on Tripadvisor or reading about a possible employer on Glassdoor. In this project, we'll discuss the history of fake review detection, examine some of the features that models use to distinguish between genuine and fake reviews and opinion spam and develop a basic fake review detection model that makes use of Tfidf Vectorizer and a number of machine learning algorithms to distinguish between the two.
The fact that phony reviews can take many different forms is one of the reasons why they might be difficult to detect. Of course, there are two basic categories of false reviews: those produced by humans and those produced by computers. However, reviews created by people as well as those created by computers can have a positive or negative tone and can be intended to enhance or lower the overall rating as well as the number of reviews to give the score more legitimacy[37]. The various false review kinds you might come across generally fit into one of the four categories below:

3.1 Computer generated reviews

Fake computer generated reviews can be produced using AI text generation models. Machine learning may be used to construct and detect these AI reviews as demonstrated by data science researchers Salminen et al. (2022)[18].

3.2 Human generated via review farms

Through review farms, which promote their services on Facebook and other websites fake reviews can be bought in volume. He, Hollenbeck and Prospero (2022) researched these phony review companies and discovered that they would be paid to create evaluations for a range of Amazon products including those with plenty of reviews and high average ratings. They discovered that businesses' share of one-star ratings considerably increased after using phony review services indicating that review manipulation was particularly common for cheap goods[22].

3.3 Human generated fake negative review

Disgruntled customers, ex-employees or rivals that wish to harm a product or company's reputation by saturating it with maliciously negative evaluations are known as human-generated false negative reviewers[22].

3.4 Human generated fake positive review

All review sites that accept them are riddled with human-generated phony positive evaluations whether they come from restaurants, Amazon marketplace sellers, e-commerce merchants or HR departments attempting to bury the Glassdoor reviews of dissatisfied former employees who have criticized the organization[22].
In-depth studies of phony reviews by researchers have uncovered a wide range of potential characteristics that both people and computer models can use to distinguish between a fake and a genuine review. Since false reviews frequently utilize similar language, especially if they are created by the same individual, business or review farm, the review text itself is typically the most crucial component[16]. To identify phony reviews, there are numerous non-text criteria that can be exploited. In Theodoros Lappas' paper from the 2012 International Conference on Application of Natural Language to Information Systems, it examines the many strategies used to avoid detection and make phony reviews appear legitimate. It is written from the attacker's point of view.
Review length sentiment: The review's word count may reveal if it's authentic or not.
Sentiment: False reviews frequently have more extreme sentiments, either being positive or extremely negative.
Helpfulness: There may be a link between phony reviews and lower helpfulness scores when review helpfulness is a platform indicator.
Reviews per user: The user who left the review might have just signed up because they made a false account with the express goal of leaving a fraudulent review because some reviews may be manufactured by bots. Models may find it advantageous to be able to detect the average amount of reviews per user.
Verified reviews: Some review sites, like Trustpilot don't demand proof that you've used a company's goods or services before leaving a review, making them vulnerable to fraudulent reviews. When a review is "verified," it means that it was written after the retailer requested it in response to a tracked customer purchase.
Stealth: "Stealth evaluates the capacity of the review to blend in with the corpus," claims Lappas (2012). Fake reviews may aim to appear like other evaluations from the company because they may stand out if they are written in a completely distinctive manner from the others.
Coherence: We may occasionally come across reviews that give a product or service a very low rating but yet include text praising it. It "evaluates whether the assigned grade is in line with the viewpoints conveyed in the review's content," according to Lappas (2012), who labels this "coherence."
Readability: Readability has been found to be a useful tool for identifying some bogus reviews by a number of machine learning researchers. For instance, Flesch Reading Ease (FRE), which other writers have also included in their models was employed by Lappas (2012).

Review text: The one element that phony review detection models use the most is the review text. Text is converted into a Bag of Words using NLP techniques like count vectorization and TF-IDF and then Naive Bayes text-classification algorithms are employed to determine whether or not the review is fake.

3.5 Fake Reviews Detector: Broad Overview

Deception variables yields accuracy levels of about 90%. The dataset that was used for the same is likewise open to the public. Additionally, ordinary n-gram text categorization methods do significantly better than human judges at detecting negative false opinion spam[37]. Multi-layer perceptrons were applied for classification in artificial neural networks' encouraging text classification performance. By examining user activity attempted to identify spamming networks using the frequency with which reviewers posted for the same items.

3.5.1 Natural Language Processing techniques

Text mining and natural language processing are terms used to describe the understanding and analysis of natural language using computer methods and algorithms. It is a prominent topic of research in the field of applications of artificial intelligence. Text mining and natural language processing studies have been studied since the invention of computers. Thanks to continuous, detailed research on machine learning and data mining algorithms, existing text mining technologies have achieved success in automated abstraction, automatic question answering, web relational network analysis and anaphora resolution[22]. An interdisciplinary field known as bioinformatics was born out of the success and advancement of the Human Genome Project. It makes predictions and resolves actual genetics-related scientific problems utilizing computer and statistical informatics. The three main steps in bioinformatics are data storage, retrieval and analysis. The National Center for Biotechnology Information developed the Online Mendelian Inheritance in Man database, the Gene Expression Omnibus database, the Sequence databases for storing DNA and protein data, the PubMed database for preserving biological and medical literature among other databases.[23].

3.5.2 Machine Learning in NLP

Artificial intelligence and machine learning have grown in popularity over the past several years. Their methodologies and ideas are now used in a vast area of products and most applications and appliances require them. The automatic identification of essential emails and speedy responses in Gmail are examples of applying machine learning. Today, we can declare with confidence that artificial intelligence and machine learning can extricate a person from many technological processes[24]. Machine learning is the scientific study of the statistical methods and algorithms that computer systems use to successfully do a certain task without using explicit instructions and instead relying on patterns and inference. It is regarded as a component of artificial intelligence. Machine learning algorithms build a mathematical model using sample data or "training data" without being explicitly instructed to

do so in order to make predictions or judgements[24, 25]. Machine learning algorithms come in five different flavors: supervised, semi-supervised, active learning, reinforcement learning and unsupervised learning. A branch of computer science, information engineering and artificial intelligence called "natural language processing" is interested in how computers interact with human (natural) languages, particularly how to teach computers to process and analyze massive amounts of natural language data.

3.5.3 Support Vector Machine (SVM) Machine Learning Model

Support-vector machine (SVM), also known as Support - Vector networks in machine learning are supervised learning models with corresponding learning algorithms that examine data used for regression and classification analysis. A well-liked machine learning method for solving classification and regression issues is the Support Vector Machine (SVM) algorithm[26]. A separating hyperplane serves as the formal definition of a Support Vector Machine (SVM), a discriminative classifier. In other words, the method generates an optimum hyperplane that classifies fresh samples given labeled training data (supervised learning). This hyperplane, which divides a plane into two portions in two-dimensional space, has one class on either side.

3.5.4 Logistic Regression Machine Learning Model

Logistic regression is most popular Machine Learning methods for two-class categorization. A statistical approach is used to forecast binary classes. - Applications like machine learning and data mining need the use of classification techniques. Used for several classification-related concerns, such as diabetes prediction and spam detection, among others[28].

3.5.5 Term Frequency Inverse Document Machine Learning Model

The preferred method for presenting documents as feature vectors is as described above. The abbreviation TF-IDF means "Term Frequency, Inverse Document Frequency." According to a document and the full corpus, a word's importance is measured using TF-IDF[30]. Authentic Review: "Lovely, bright, and tidy room. Excellent views, peaceful, a decent bed, etc. Excellent staff and overall top-notch service. For the money, the quality and value were superb. Also at a great location. False rating: "Dark and disorganized room. Loud, uncomfortable beds, poor views, etc. Poor overall service and unfriendly staff. It was of very low value and quality, and the location was awful. The review sample stated above amply demonstrates how spammers reverse the polarity of aspects like room, bed, staff, and location. Tags like stunning, clean, and bright might be used to infer the atmosphere of the place. The polarity of the room is altered through the use of phony evaluational words like untidy and dark. This sample was picked from many line evaluations in order to better demonstrate our technique. The aforementioned example demonstrates that it is unnecessary to focus on in-depth text analyses of several lines and words because doing so will just make the computing process take longer[15,16]. The review

sample stated above amply demonstrates how spammers reverse the polarity of aspects like room, bed, staff, and location. Tags like stunning, clean, and bright might be used to infer the atmosphere of the place. The polarity of the room is altered through the use of phony evaluational words like untidy and dark. This sample was picked from many line evaluations in order to better demonstrate our technique. The aforementioned illustration demonstrates that it is unnecessary to focus on comprehensive text evaluations of several lines and words because doing so will just make the computation process take longer[15, 16]. Because spammers are unfamiliar with the product, they change a few words to create fictitious reviews. The authors have discovered that the analysis of fake reviews can be done using part of speech (POS) tagging. Our approach uses POS tagging to tag nouns and noun phrases as aspects, adjectives, adverbs, and verbs as sentiments. Deep learning algorithms outperform traditional classifiers when assessing spam comments. The reason for this is that traditional classifiers lose their ability to provide precision once they achieve a certain level of accuracy. Their suggested approach employs a hybrid CNN and LSTM model for aspect replication and sentiment learning. Extracted aspects and their respective polarities are fed into a CNN model in order to find aspect repetition. Then, for training and performance evaluation, the filtered aspect replication is fed into an LSTM[15].The following are the primary contributions of this research work: 1. Aspect and polarity from reviews are extracted using eective POS tagging-based algorithms. 2. Their method is superior to existing approaches because it computes aspect replication in spam reviews rather than whole text replication. 3. In order to analyze bogus reviews, extracted characteristics and feelings are fed into deep learning models (RNN and LSTM). 4. Ott and Yelp filter datasets are used frequently in experiments to analyze accuracy. Analysis of the results of experiments demonstrates that our suggested strategy oers greater precision and accuracy when compared to current approaches. Aspect extraction and spam opinion approaches that are related to this study are reviewed. Background information and preliminary steps are provided for aspect extraction methods, RNN, LSTM architecture and spam opinion detection. The suggested strategy is explained. The design and analysis of experiments are discussed[16]. To identify fraudulent reviews, previous research has concentrated on the reviewer's behavior or the review wording. The most typical technique employed by spammers to produce phony reviews is the usage of sentiments and material that has been taken directly from reviews. This part surveys current state-of-the-art methods to evaluate the benefits and drawbacks of previous research, which inspires us to suggest a practical method for identifying false reviews with increased precision and reduced computing time.

3.6 Analysis of fake reviews using review text

The use of neural networks to identify false opinions using document-level representations is investigated. The authors of this work state that earlier research only paid attention to discrete qualities that were based on linguistic viewpoints. Discourse and document-level semantics are taken into consideration in this research project. In order to analyze the discourse semantics, a recurrent neural network receives input from the CNN model, which is used to learn representation from phrases. Three datasets were used for the experiments and it was determined that the proposed strategy enhanced accuracy. AutoEncoder (DAE) and Vector Distributed Bag of

Words (PVBOW) is suggested. The pre-processing of the dataset uses lemmatization and tokenization. For hidden layers, ReLU is utilized as the activation function, and Sigmoid function is used for output layers. For the analysis of the experiment, gold standard dataset is employed. When evaluating the suggested strategy, accuracy, F1-measure, precision and recall are taken into consideration. Reviewers are given a reputation score to determine whether they are legitimate or spam. This method has the benefit of not requiring huge instances to be labeled[18]. The application of additional k-centre clustering depends on time interval. Reviews posted by spammers allegedly have significant emotional inclinations. Combining content attributes with reviewer behaviors yields reputation value. Amazon music product reviews serve as the dataset for the experiment analysis. In this study, precision, recall and F-measure are used. In comparison to current methodologies, the performance of the proposed methodology is better. For experiment analysis in this paper, a dataset of Amazon product reviews is employed, together with the entire review content. To extract aspects, a deep convolution neural network is used. Seven-layer neural networks have been utilized by the authors to increase aspect extraction's precision. It is claimed that linguistic patterns and conditional random fields have shortcomings that need to be fixed. For experiment analysis, datasets from Google, Amazon and SemEval are used. The proposed technique outperforms LP when precision values and language patterns are examined. There are several neural network designs that are explicitly discussed for spam opinions. This study work also makes the claim that conventional machine learning algorithms do not offer the semantic information of reviews that is required for the analysis of misleading spam opinions. Numerous neural network topologies including CNN, RNN, LSTM and GRU are used in experiments. In terms of accuracy, CNN outperforms other models. It's because CNN can extract sophisticated and intricate features from user opinions. Information from the review text is extracted using the N-Gram model. The ensemble method which combines CNN and N-gram is applied. On the Yelp dataset, experiments are run to evaluate the effectiveness of the suggested approach. The drawback of this method is that it relies on a number of reviewer characteristics and reviewer behavior to identify phony reviews. The computational complexity rises as a result. For the purpose of detecting false reviews, various machine learning and deep learning models are employed. On the Yelp and Ott datasets, the performance of CNN, LSTM, SVM and kNN models is assessed. The drawbacks of this strategy include the usage of a single model for the detection of false reviews and the possibility of word embedding and hyperparameter settings enhancement. Utilizing semi-supervised learning, bogus reviews can be found. The authors claim that while precise labels and extensive data are needed, labeled datasets can present a problem. In order to choose features that can perform better in less computing time, this approach needs optimization[19].

3.7 Analysis of fake reviews using reviewer behavior

Entities are suggested as a method for detecting opinion spam. In this method, entities are given weights based on how important they are. This study discusses spam characteristics including review polarity intensity, words, rating deviation, re-

views per day, etc. Utilizing the evaluation criteria Accuracy, Precision, Recall, and F-measure, the suggested approach is verified. This study uses a range of characteristics, including content-based, behavior-based, relation-based, and proposed features, to measure correctness. Accuracy is improved by using the indicated set of qualities. The authors have proposed a method based on singleton reviews. Few research, according to the authors, have focused on reviewers who have only published one review. It is possible for reviewers to submit spam reviews by changing their usernames. Between reviews, textual and semantic similarity is computed. On the datasets from Yelp and Trustpilot, experiment analysis is done. By utilizing the suggested approach, precision and F1 score are increased. To identify spam opinions, autoencoder and neural random forest are utilized. These models were chosen because random forest combines many decision trees while autoencoder can employ unsupervised representations in features. This study took into account both reviewer behavior and review substance. Features include things like the summary's word count, the ratings' entropy, when they were given, etc. The experiment analysis uses the Amazon review dataset, and the evaluation metrics used are precision, recall, and F1-measure. When compared to the current approach, it has been shown that the proposed approach offers more accuracy. The computational challenge is increased by the use of the complete review text for semantic analysis. From content- and behavior-based aspects, 133 traits are gathered. The class imbalance in the datasets is mentioned. To resolve this, a random sampling is employed. Accuracy is improved by sampling[20]. This approach does not, however, use many classifiers and there is room for development to reach satisfactory accuracy on sizable datasets. Reviewer behavior is examined in order to identify bogus reviews. Star User, Deviation Rate, Bias Rate, Review Similarity Rate, Review Relevancy Rate, Content Length, and Illustration are a few of these. In order to improve the classification, sophisticated neural network-based models must be used. The novelty, advantages, disadvantages, assessment criteria and dataset of contemporary systems to fake review identification are analyzed. Our analysis reveals that whole text reviews are employed in previous studies which raises the computational di culty. Reducing computational complexity is necessary. Additionally, shallow architectures are employed for the detection of false reviews, necessitating the optimization of the neural network design through the use of dropout, e cient feature selection and hyperparameter tuning. Aspects are not given significant attention in current study activities. Only pertinent features are used in our suggested method, which reduces computing complexity. Iterations of the complexity analysis are made. Additionally, the majority of contemporary approaches use deep learning or conventional machine learning models. In our suggested method, a hybrid deep learning model is used to take use of the benefits of CNN and LSTM as mentioned. Recent research has shown a significant increase in interest in the use of NLP for opinion spam identification and false review detection. Due to this, a large number of literature reviews and research articles have been written with the specific goal of identifying fraudulent reviews or opinion spam. The emphasis of this section is on showcasing earlier research in this area. Classification models may help distinguish between reviews and classify them as either real or fake, which makes them quite useful when it comes to applying machine learning to identify fraudulent reviews. A dataset of 2.14 million reviews, many of which were duplicates, was utilized to generate classification predictions for reviews using the conventional machine learning technique of logistic

regression. A rudimentary framework for spotting spam reviews was given using a combination of supervised, semi-supervised, and unsupervised learning. For the purpose of detecting fraudulent reviews, a thorough approach that included data gathering, data pre-processing, and machine learning was offered. Semantic similarity between terms at the review level and topic modeling were employed in an effort to spot fake reviews that had been written by the same person but published under various names. The popular Yelp algorithm for identifying fake restaurant reviews performed well for behavioral features but not so well for linguistic features, it was also found. Singular Vector Machines (SVM) have been shown to perform better in text classification than other traditional data mining algorithms[19]. It is also possible to tell whether a review is phony or real by counting how many users have given it a helpful vote or upvote, which measures how beneficial the review is. Additionally, it was discovered that combining a classifier with both n-gram and psychological over short periods of time. They also took into account the frequency with which other users posted for the same products during those same periods of time. For topic modeling, Probabilistic Latent Semantic Analysis (pLSA) can be used to automatically extract aspects and organize them into dierent categories. Latent Dirichlet Allocation (LDA), a generative probabilistic model that assumes that each document is built of a variety of themes and that each topic is a probability distribution over words, can also be used for topic modeling, which was introduced as a technique for identifying phony reviews. For text mining, three additional feature categories were introduced: review density, semantic similarity, and emotion. The review centric approach and the reviewer centric strategy were presented as an overview of the current detection techniques.

Chapter 4

Methodology

4.1 Workflow

In our little time of research, we believe that the workflow is the most crucial component of a study. Our work cycle is outlined in Figure 4.1. We used this approach in order to perform as intended or to maximize the outcome. The initial step in our approach was choosing the data sets. The dataset was then preprocessed. In data preprocessing we removed the punctuations and stopwords. Then, we replace the target value. Then we split the data where we split the value as train, test and val. Then we jump in the next part word embedding. In word embedding part we use tfidf and glove. Then we will use Support Vector Machine (SVM) and Logistic Regression (LR) in Glove. Moreover, to do the train test part we will use the value of glove from Logistic Regression and Support Vector Machine Model. By this way we get the accuracy. By using this accuracy we will find do the performance train.

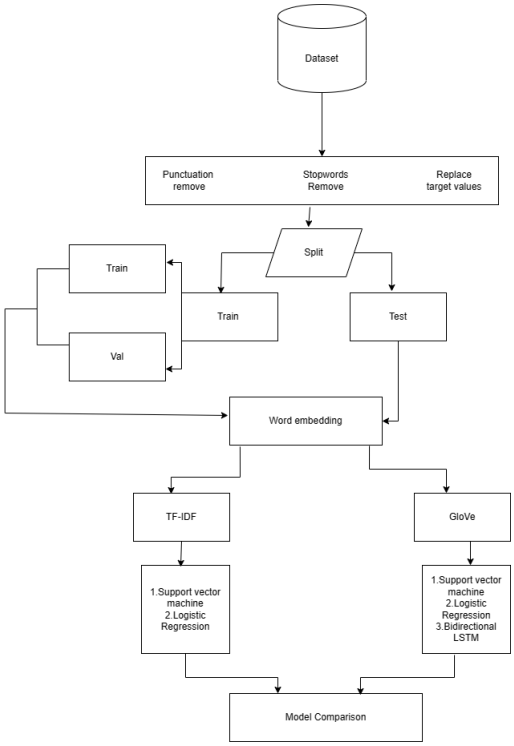


Figure 4.1: Workflow

4.2 Dataset

The dataset we used in this study has been taken from OSF — Pdfs from dot . This dataset contains 40,431 customer reviews of different products. Here we have 2 different types of reviews as in computer generated reviews and original reviews from people. Computer generated reviews are labeled as CG and original reviews from people are labeled as OR in the dataset.
We faced difficulty while searching for our desired dataset. At first we chose a dataset from Amazon which contained approximately 14 lakhs reviews which was quite large for our research. Most of the datasets available online are more or less of the same size and quite large and robust. For this reason, we had to surf the web well to finally find our desired dataset. A proper dataset is needed because using raw data companies may establish baselines, benchmarks and goals in order to advance. Data enables measurement allowing you to establish baselines. A baseline is the condition of a place before a particular remedy is applied.

4.3 Preprocessing

In order for businesses to undertake data mining, evaluate the data and process it for business activities, raw data must be transformed into legible and defined sets. Preprocessing by deleting missing or inconsistent data values caused by human or machine mistake data may improve the accuracy and quality of a dataset making it more dependable. Preprocessing raw data effectively can boost its precision, which will boost projects' quality and dependability. Data duplication is a possibility while collecting data, and removing them during preprocessing can guarantee that the data values for analysis are consistent, assisting in the production of reliable results. At first we splitted the dataset into two groups where 90% of the dataset was reserved for training the model and 10% was kept for testing. Then again we splitted the training dataset into two groups where 90% of the dataset was assigned for training the model and 10% was kept for validation. As a result, we got 81% of the main dataset as training set, 10% as testing set and 9% as validation set. We kept 9% of the dataset as validation for tuning the dataset.
Data is transformed into discrete variables as the initial step in dataset pre-processing. To make things function, we had to turn the attribute domains into discrete variables. Machine learning models cannot utilize text on their own. They anticipate receiving numerical input.
Data engineering and feature engineering are both involved in the preprocessing of the data for ML. The prepared data is subsequently tuned through feature engineering to provide the features that the ML model anticipates.
Here we converted every text of our dataset to lowercase as this dataset will be used as our main input. For output we will need OR and CG labeled data but as machine learning models cannot use text on their own we replaced 'OR' as 0 and 'CG' as 1. For this, now we will get output as 0 or 1. Then we removed all the punctuation marks from our dataset and created a new column named twp. Then from this column we removed all the stop words(e.g "a", "the", "is", "are" and etc.) and created another column named tws. For further operation we used this tws column.

4.4 Model Implementation

We examined several supervised classification algorithms for this work and Support Vector Machines (SVM), a well-known technique that has been successful in many areas, including deception detection which produced the best results. Its ability to cope with things that would not be linearly separable in the feature space, using kernel functions that move the entities in a higher dimension space, where the linear separation is achievable is what determines how successful it will be. As a result, the kernel function selection is essential to the models' efficacy. Although texts are typically represented in vector space by sparse vectors, radial kernels performed the best in our trials even though linear kernels are typically thought to be effective for text categorization. Because of the relatively modest size of our corpus, which would have prohibited the training of accurate word embeddings relevant to the task, we chose to employ SVMs rather than the Deep Learning techniques that have been frequently used in recent literature on text categorization. The majority of the literature on deception detection uses more conventional classifiers, which allows for a more direct comparison of our results to other researchers' findings. Tenfold cross-validation was used as the validation approach during the model's training. This paper's evaluation of the widespread practice of generating synthetic data sets for the identification of false reviews via crowdsourcing was one of its primary objectives. This approach is frequently employed in the literature, which typically reports positive outcomes. Our findings, however, indicate that this strategy might not be the best one. Additionally, we observe a significant change in the prediction accuracy depending on how close the training and test sets are in terms of the domain. For model implementation, first of all we built the model and installed the packages. Then, we imported the packages and loaded all the data. If there is any data imbalance then checked the data imbalance. After preparing the data, we created features from punctuation, tokenized the data and used stopword removal. Then after applying porter stemming and rejoining words we created training testing data and run the model selection process. By this we accessed our proposed model.

4.5 TF-IDF

The metric TF-IDF stands for frequency-inverse document. Frequency can be used to assess how important or pertinent string representations (words, sentences, lemmas, etc.) are inside a document in comparison to other texts. It is utilized in the fields of machine learning and information retrieval (IR). In TF-IDF there are two parts: TF (Term frequency) and IDF (Inverse document frequency).
The TF-IDF states that a term's value is inversely correlated with how frequently it appears in documents. While TF offers information about how frequently a phrase appears in a document, IDF provides information about the relative rarity of a term in the collection of documents. By averaging these figures, we might arrive at our ultimate TF-IDF value. The following mathematical formulas are used to determine the TF-IDF score for the word t in the document d from the document collection D:

$$tfidf(t, d, D) = tf(t, d).idf(t, D) \tag{4.1}$$

In Tfidf vectorizer we applied fit transform on our trained dataset and applied only transform on validation and test dataset.

4.6 GloVe Vectorizer

Global vectors for word representation is referred to as GloVe. The global word-word co-occurrence matrix from a corpus is combined with this Stanford-developed unsupervised learning technique to create word embeddings. The resulting embeddings in vector space emphasize the word's intriguing linear substructures. In our proposed framework, we took the txt file of GloVe Vectorizer from Stanford website. For every word it will have 100 floating points and the file had approximately 4 lacs floating points for different words. We took this into our dataset then we converted it to dictionary. Then we calculated floating point value for each word of every review from our dataset and stored them to a list. We added all the floating point value of every word of a single review and divided it by the total number of words and stored it in our dataset. After that we use this in SVM and LR as input. But for LSTM we did not add the floating point of every word for a sentence. Here we took floating value for every word and append the value in a 3d list. In that 3D list every row represents a review. Then we converted that 3d list into dataset and use that dataset for input in LSTM model.

4.7 Support Vector Machine Algorithm(SVM)

Currently the majority of machine learning tasks involve classifying images, translating languages, managing vast quantities of sensor data and making predictions about the future value based on the present.For both classification and regression issues, Support Vector Machine (SVM), amongst the most well-liked supervised learning techniques, have been used. The SVM method seeks to define the best line or decision boundary which can divide n-dimensional spaces into classes in order to quickly categorize new data points in the future. This boundary's best option is known as a hyperplane.As they select the decision boundary that optimizes the distance from the nearest data points of all the classes, SVMs vary from other classification techniques. How SVM functions is discussed briefly below -
i) Linear Case - The problem of two classes with N training samples should now be taken into consideration. A Support Vector (SV) Xi comprised of various "bands" with n dimensions is used to describe each sample. Yi is written on a sample's label. In the event of two classes, we take into account the labels 1 for the first class and +1 for the second. The function is defined by the SVM classifier -

$$f(x) = sign(<\omega, X> + b) \tag{4.2}$$

which determines the ideal separating hyperplane , where ω is the hyperplane's normal and $\frac{|b|}{||\omega||}$ is the hyperplane's distance from the origin perpendicularly. The label of the sample is provided by the sign of f(x). The SVM seeks to maximize the distance between the support vector and the ideal hyperplane. Thus, we search the min $\frac{|\omega|}{2}$

Using the Lagrange multiplier makes this process simpler. The issue has to be resolved:

$$f(x) = sign(\sum_{i=1}^{Ns} .y_i.\alpha_i<x.x_i> + b) \tag{4.3}$$

where α_i is the Lagrange multiplier
ii) Nonlinear Case -
In our proposed framework, we used four kernels (rbf, linear, poly, sigmoid). Data can be entered and then changed into the format required for processing using a kernel function. The word "kernel" is used because a Support Vector Machine's window for data manipulation is given by a series of mathematical operations.
Three kernels are commonly used: The polynomial kernel:

$$K(x, x_i) = (<x, x_i> + 1)^p \tag{4.4}$$

The sigmoid kernel:

$$K(x, x_i) = tanh(<x, x_i> + 1) \tag{4.5}$$

The RBF kernel:

$$K(x, x_i) = exp^{\frac{|x-x_i|^2}{2\sigma^2}} \tag{4.6}$$

We got different accuracy scores for each of them in different regularization parameter(CS = 1.0,1.1,1.2). We got the highest accuracy score for rbf which is 89.86%.

4.8 Logistic Regression(LR)

Another type of supervised learning technique is logistic regression. Binary logistic regression is the most often used application of logistic regression where the outcome is a binary decision (yes or no). A logistic regression model forecasts the value of P(Y=1) as a function of X mathematically.It may be applied to many categorization issues, such as spam detection, diabetes prediction, cancer diagnosis, etc.
Here is an illustration of a logistic regression formula:

$$y = e^{\frac{b_o+b1*x}{1+e^(b_o+b1*x)}} \tag{4.7}$$

If we take b0 is the bias/intercept term, then b1 = coefficient of the single input value (x) and y = the anticipated output. You must learn the associated b coefficients (constant real values) for each column in your input data from your training set.
In our proposed framework we used different solvers(newton-cg, lbfgs, liblinear, sag, saga).We got different accuracy scores for each of them in different regularization parameter(CS = 1.0,1.1,1.2). Each observation gives an equal amount of information, which is a key premise of logistic regression. There is a chance to give a Weight variable while using Analytic Solver Data Mining. The user can assign a weight to each entry by using the Weight variable. We got the highest accuracy score for newton-cg and sag in this case which is 87.94%.

4.9 BiDirectional LSTM

Even though we got a decent result with Logistic Regression and Support Vector Machine, we aimed to improve the accuracy of our results by utilizing the power of neural networks. We chose to work with Recurrent Neural Networks (RNNs) as they have a proven track record in working with sequential data. However, due to the long-term dependency problems of traditional RNNs, we decided to implement Long Short-Term Memory (LSTM) networks.
Initially, a single-directional LSTM with Glove Vectorizers provided us with an accuracy of 85%. However, this performance was not as high as we had hoped. Therefore, we attempted to increase the accuracy further by implementing a Bidirectional LSTM model. The model consisted of one Global Max pooling 1-dimensional layer, followed by a batch normalization layer to speed up the training process with higher learning rates. Additionally, there were three back-to-back dropout layers, followed by one dense layer each, with the first two having relu and the last having sigmoid as the activation function. The dropout layers were crucial in preventing overfitting by shutting down the contributions of certain neurons. The model had an input dimension of (237 x 100). To compile the model, we used Root Mean Squared Propagation (RMSProp) optimizer to accelerate the Gradient Descent process. Compared to AdaGrad, RMSProp performed faster as it utilizes the decaying average of the partial derivatives. The loss was calculated using the Binary Cross-Entropy loss function, as it is best suited for datasets with 2 target classes.

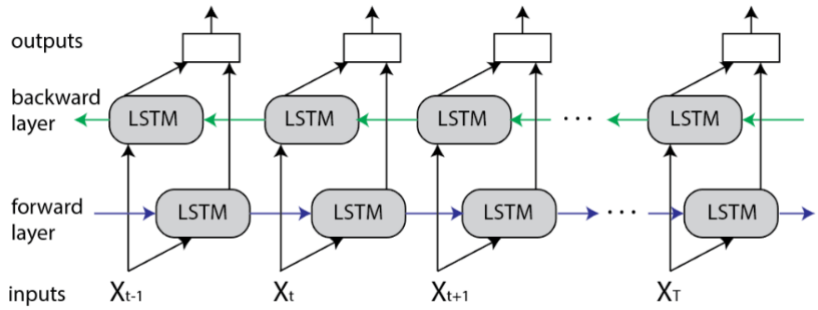


Figure 4.2: Bidirectional LSTM

4.10 Confusion matrix, Accuracy, F-1 Score , Recall and Prescision

A confusion matrix is a table that shows the results of a classification problem, including the predicted and actual values of a classifier. It is used to visualize the outcomes of the prediction and helps to understand the performance of the classifier. There are 4 types of values in confusion matrix.

$$1)TruePositive \tag{4.8}$$

$$2)TrueNegative \tag{4.9}$$

$$3) False Positive \tag{4.10}$$

$$4) False Negative \tag{4.11}$$

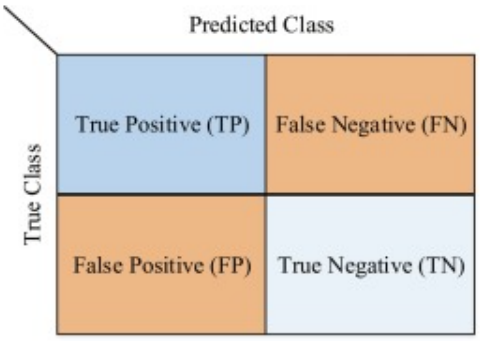


Figure 4.3: Confusion matrix

A measure of accuracy means the proportion of correctly classified data cases to the total number of data cases, which indicates how many data cases were correctly classified.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \tag{4.12}$$

The precision of a classifier's accurate positive predictions is measured. It is defined as the proportion of accurate positive predictions to all of the classifier's positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{4.13}$$

Recall is a metric for how many out of all the positive cases in the data that the classifier correctly predicted.

$$Recall = \frac{TP}{TP + FN} \tag{4.14}$$

The F1-Score is a statistic that combines a classifier's precision and recall. It is computed as the harmonic mean of recall and precision, which is a separate method than calculating the variables' "average." The regular arithmetic mean is seen to be less suitable for ratios like precision and memory than the harmonic mean.

$$F - 1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4.15}$$

Chapter 5

Experiments and Results

5.1 Result analysis

The dataset for this study was collected from OSF — Pdfs from dot. 40,431 consumer reviews of different products are included in this dataset. Here, there are two different kinds of reviews: those produced by computers and those that real people have written. In the dataset, computer produced reviews are denoted as CG and manually authored reviews as OR. There are total 20216 reviews which is computer generated and other 20215 reviews which is real reviews from people. In our datasets there we used two columns one is text and another one is label where text column represents all the reviews and label represents if the review is CG(Computer Generated) or OR(Original Review). At first, we splitted our dataset in two set one is for train set where the test size is 10% of the main dataset and train set is 90% then again we splitted the train set into train set and validation set where train set has 81% data of the main dataset and validation set has 9% data. We splitted twice one is for train test another one is for train validation. The second split we did for tuning our model. Model tuning is a process to discover the optimal values in order to maximize the performance of the model. So after that we implemented TF-IDF vectorizer on input part which are x-train,x-test and x-val. Next we also implemented GloVe vectorizer on our input part of the dataset. Here we have used Support Vector Machine(SVM). Logistic Regression(LR) and LSTM. We used SVM and LR two times one with TF-IDF vectorizer value and other one with GloVe vectorizer value.

5.1.1 SVM and LR Results using TFIDF:

First we used a loop to try different types of parameters in our SVM model. In parameters we set four types of kernel ('rbf', 'linear', 'poly', 'sigmoid') then we set max iteration -1 and we used different regularization parameter(CS = 1.0, 1.1, 1.2). We checked accuracy with each CS for each kernel. We got the best accuracy (89.50%) and weighted f1 score(0.84) when kernel = "rbf" and c=1.2 and we got less accuracy(84.36%) and weighted f1 score(0.84) when kernel = "poly" and CS= 1.0. . Then we used train set and test set in SVM model by setting the value of kernel = "rbf" and c=1.2 for these values we got weighted f1 score 0.90 and accuracy 90.31%. Then we did the same thing using LR but here instead of using kernels we used solvers(newton-cg, lbfgs, liblinear, sag, saga). Then we set max iteration

500 and we used different regularization parameter(CS = 1.0, 1.1, 1.2) here also. We checked accuracy with each CS for each solver. Then we used train set and validation set in LR model by setting different values of solver and CS. For solver = "newton-cg" and CS=1.2 we got weighted f1 score 0.88 and accuracy 87.94%. We got the best accuracy in this case. We got less accuracy(87.80%) and weighted f1 score(0.88) when kernel = "sag" and CS= 1.0. Then we used train set and test set in LR model by setting the value of For solver = "newton-cg" and CS=1.2 for these values we got weighted f1 score 0.89 and accuracy 88.92%.

Model	Precision	Recall	f1-Score	Accuracy
RBF 1.0	0.92	0.87	0.89	0.90
RBF 1.1	0.91	0.88	0.89	0.90
RBF 1.2	0.79	0.88	0.89	0.90
linear 1.0	0.89	0.87	0.88	0.88
linear 1.1	0.89	0.87	0.88	0.88
linear 1.2	0.89	0.87	0.88	0.88
poly 1.0	0.94	0.74	0.83	0.84
poly 1.1	0.94	0.75	0.83	0.85
poly 1.2	0.94	0.76	0.84	0.85
sigmoid 1.0	0.89	0.85	0.87	0.87
sigmoid 1.1	0.88	0.85	0.87	0.87
sigmoid 1.2	0.88	0.85	0.87	0.87

Table 5.1: Results of using different parameters of SVM with TFIDF

Model	Precision	Recall	f1-score	Accuracy
newton-cg 1.0	0.90	0.85	0.88	0.87
newton-cg 1.1	0.90	0.86	0.88	0.87
newton-cg 1.2	0.90	0.86	0.88	0.87
Lbfgs 1.0	0.90	0.85	0.88	0.87
Lbfgs 1.1	0.90	0.86	0.88	0.87
Lbfgs 1.2	0.90	0.86	0.88	0.87
Liblinear 1.0	0.90	0.85	0.88	0.87
Liblinear 1.1	0.90	0.86	0.88	0.87
Liblinear 1.2	0.90	0.85	0.88	0.87
Sag 1.0	0.90	0.86	0.88	0.87
Sag 1.1	0.90	0.85	0.88	0.87
Sag 1.2	0.90	0.86	0.88	0.87
Saga 1.0	0.90	0.85	0.88	0.87
Saga 1.1	0.90	0.86	0.88	0.87
Saga 1.2	0.90	0.85	0.88	0.87

Table 5.2: Results of using different parameters of LR with TFIDF

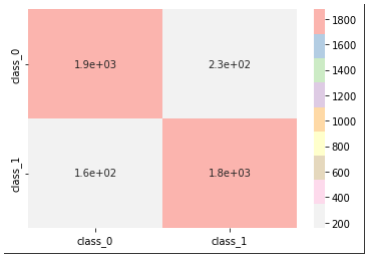


Figure 5.1: Confusion matrix for SVM usign GloVe

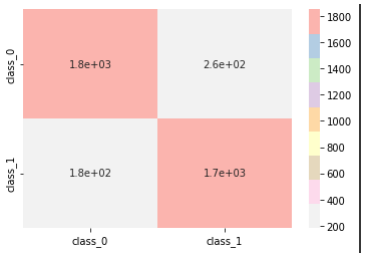


Figure 5.2: Confusion matrix for LR usign GloVe

5.1.2 SVM and LR Results using GloVe:

Secondly for GloVe vectorizer we got best accuracy=79.44% and weighted f1 score= 0.79 using SVM. We got less accuracy 45.07% and weighted f1 score= 0.45 when kernel = "sigmoid" and CS= 1.1. Then for train and test dataset we used kernel = "rbf" and CS = 1.2 as we got best accuracy for these values. After applying it we got 0.8 as f1 score and 80.34% as accuracy. Then we did the same thing for LR. We got best accuracy=76.17% and weighted f1 score= 0.76 using LR when solver = "lbfgs" and CS=1.0. We got less accuracy 76.12% and weighted f1 score= 0.76 when solver = "newton-cg" and CS= 1.1. Then for train and test dataset we used solver = "lbfgs" and CS = 1.0. After applying it we got 0.77 as f1 score and 77% as accuracy.

Model	Precision	Recall	f1-Score	Accuracy
RBF 1.0	0.79	0.79	0.79	0.79
RBF 1.1	0.78	0.81	0.80	0.79
RBF 1.2	0.79	0.82	0.80	0.79
linear 1.0	0.77	0.76	0.76	0.76
linear 1.1	0.77	0.76	0.76	0.76
linear 1.2	0.77	0.76	0.76	0.76
poly 1.0	0.80	0.77	0.78	0.79
poly 1.1	0.80	0.77	0.78	0.79
poly 1.2	0.80	0.77	0.78	0.79
sigmoid 1.0	0.46	0.44	0.45	0.45
sigmoid 1.1	0.45	0.44	0.45	0.45
sigmoid 1.2	0.45	0.44	0.45	0.45

Table 5.3: Results of using different parameters of SVM with GloVe

5.1.3 Bidirectional LSTM accuracy and Loss function:

But for LSTM after 7 epochs, the model achieved an accuracy of 0.9166 and a binary cross-entropy loss of 0.1533. This was the best score we found among all the models we tested, although we believe that it can be further improved with additional tuning and training.

Model	Precision	Recall	f1-score	Accuracy
newton-cg 1.0	0.77	0.76	0.76	0.76
newton-cg 1.1	0.77	0.76	0.76	0.76
newton-cg1.2	0.77	0.76	0.76	0.76
Lbfgs 1.0	0.77	0.76	0.76	0.76
Lbfgs 1.0	0.77	0.76	0.76	0.76
Lbfgs 1.1	0.77	0.76	0.76	0.76
Lbfgs 1.2	0.77	0.76	0.76	0.76
Liblinear 1.0	0.77	0.76	0.76	0.76
Liblinear 1.1	0.77	0.76	0.76	0.76
Liblinear 1.2	0.77	0.76	0.76	0.76
Sag 1.0	0.77	0.76	0.76	0.76
Sag 1.1	0.77	0.76	0.76	0.76
Sag 1.2	0.76	0.76	0.76	0.76
Saga 1.0	0.77	0.76	0.76	0.76
Saga 1.1	0.77	0.76	0.76	0.76
Saga 1.2	0.77	0.76	0.76	0.76

Table 5.4: Results of using different parameters of LR with GloVe

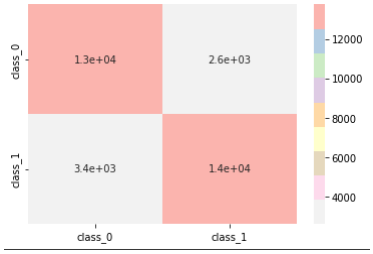


Figure 5.3: Confusion matrix for SVM usign GloVe

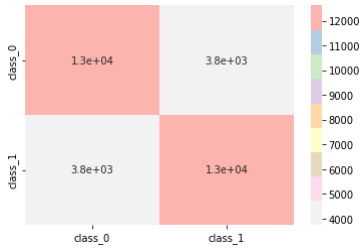


Figure 5.4: Confusion matrix for LR usign GloVe

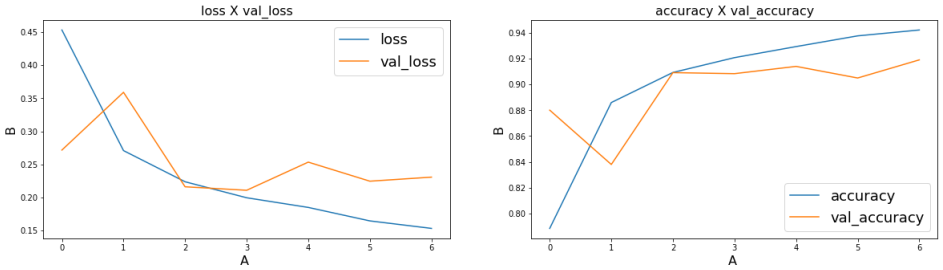


Figure 5.5: Bidirectional LSTM Accuracy and Loss Function

5.1.4 Discussion:

Word embeddings may be a preferable option for most tasks where tf-idf is employed, especially where the work may benefit from the semantic similarity recorded by word embeddings, even if tf-idf is a straightforward scoring method and that is its main advantage. If GloVe is trained properly it captures all the meaning. In NLP task GloVe should carry more information than Tf-Idf. Because it gives us 100 value for

a word whereas tfidf assign only one value based on how important the word is for this document. Here,Glove didn't perform well because here we took Glove value for each word and add that value with others words of the review and then divided that value with the number of words that review has so we got an average value for a review. So SVM and LR might not catch the proper info while training and that caused for the less accuracy. Besides, we got better accuracy for Bidirectional LSTM because here we used GloVe value for each word and directly pass the value to the model. From our analysis we got accuracy 90.30 percent for SVM and 88.92 percent for Logistic Regression where we used TF-IDF and For Glove vectorizer we got accuracy 80.34 percent for SVM and 77.00 percent for Logistic Regression. In our proposed architecture we got more accuracy using Tf-Idf than GloVe for svm and LR. One of the reasons why this happened maybe GloVe was not able to capture more information than Tf-Idf which could have resulted in underfitting the model. As the input vector size in Tf-Idf was bigger than GloVe, it might have been underfitted because of having less features than Tf-Idf. But for LSTM GLove result did better result.

Number	Model	Accuracy
1	Bidirectional LSTM	0.9166
2	SVM for TF-IDF	0.9030
3	Logistic Regression for TF-IDF	0.8892
4	SVM for GloVe	0.80341246
5	Logistic Regression for GloVe	0.7700

Table 5.5: Models accuracy comparison table

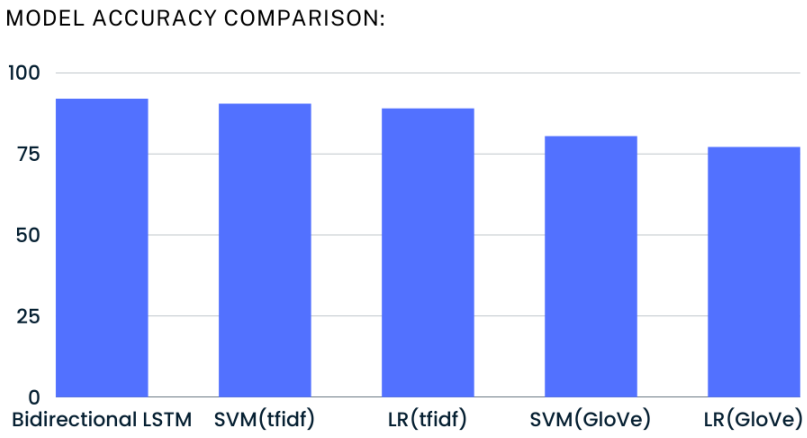


Figure 5.6: Models Accuracy Comparison

Chapter 6

Concluding Remarks and Future Work

6.1 Conclusion

Here, we'll aim to investigate the possibilities for spotting bogus reviews that are based on actual data by integrating active learning. In this work, various indicators for spotting phony reviews are proposed and they are all given weights using active learning. In the process, we train our model using active learning which iteratively learns from the best data. The review content's TF-IDF values will be used to build the feature vectors and classifiers like SVM and LR will be employed in the classification process. Our experimental methods examine the precision of each sentiment categorization algorithm and identify the technique that is most precise. Additionally, by using detection techniques, we were able to identify phony negative and good reviews. The bogus reviews are filtered using the unfair review detection method. In this study, the SVM classifier outperformed the Logistic Regression in terms of classification accuracy. Demonstrating how effectively it predicts false reviews examining the differences in classifier accuracy was made easier by the data visualization. Identifying false reviews is a difficult task. The absence of the tagged dataset is the primary problem in this particular field of study. We suggested a supervised method that does not require label information to predict the review class in order to close this gap (fake or real). Research to identify unfair reviews is extremely valuable in ensuring the reliability of reviews and giving customers a positive purchasing experience. Makers can obtain accurate data by using this strategy for detecting unjust reviews. By examining how customers feel about things, businesses may monitor their product sales and reach. Customers have the option of making a purchase or not. This technique so increases the credibility of e-commerce websites. It has been noted that fake reviews are difficult to detect throughout this study. Numerous studies have been conducted on this subject, but none of them have produced a perfect conclusion (one hundred percent result). There are still many flaws that are not being fixed, even in this day and age. We provided a way for determining whether the provided comments on a specific good or service are genuine or false using machine learning-based. Extending the dataset currently in use and determining the best outcomes for a big quantity of data will offer insights on performance in terms of accuracy as well as scalability. Also in future we will try to give less value in dropout layer of BiDirectional LSTM and try to run more

epochs for better accuracy.

6.2 Limitations and Future Work

As we run our code in Google Colab free version so we could only run our code with GPU for 5 hours constatntly after that it showed us the runtime session ended. So we couldn't run our LSTM model for long time which was needed. So in future we will try to use Colab pro in future and get the better accuracy. Also for SVM and LR using GloVe embedding we got less accuracy than TFIDF which shouln't be the case but we got this accuracy because we used average GloVe value for a sentence words so we will try to use GloVe vectorizer with better approach which will give us the better accuracy than TFIDF word embedding.

Bibliography

[1] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 92–100.

[2] A. Z. Broder, "Identifying and filtering near-duplicate documents," in *Annual symposium on combinatorial pattern matching*, Springer, 2000, pp. 1–10.

[3] O. Chapelle, B. Schölkopf, A. Zien, *et al.*, "Semi-supervised learning, vol. 2," *MIT Press, Cambridge. Cortes C, and Mohri M, et al.(2014) Domain adaptation and sample bias correction theory and algorithm for regression. Theoretical Computer Science*, vol. 519, p. 103 126, 2006.

[4] B. Stein, M. Koppel, and E. Stamatatos, "Plagiarism analysis, authorship identification, and near-duplicate detection pan'07," in *ACM SIGIR Forum*, ACM New York, NY, USA, vol. 41, 2007, pp. 68–71.

[5] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the 2008 international conference on web search and data mining*, 2008, pp. 219–230.

[6] F. H. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," in *Twenty-second international joint conference on artificial intelligence*, 2011.

[7] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," *arXiv preprint arXiv:1107.4557*, 2011.

[8] R. Y. Lau, S. Liao, R. C.-W. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," *ACM Transactions on Management Information Systems (TMIS)*, vol. 2, no. 4, pp. 1–30, 2012.

[9] A. Morales, H. Sun, and X. Yan, "Synthetic review spamming and defense," in *Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 155–156.

[10] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?" In *Proceedings of the international AAAI conference on web and social media*, vol. 7, 2013.

[11] S. Shojaee, M. A. A. Murad, A. B. Azman, N. M. Sharef, and S. Nadali, "Detecting deceptive reviews using lexical and syntactic features," in *2013 13th International Conference on Intellient Systems Design and Applications*, IEEE, 2013, pp. 53–58.

[12] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, pp. 1566–1576.

[13] M. I. Ahsan, T. Nahian, A. A. Kafi, M. I. Hossain, and F. M. Shah, "Review spam detection using active learning," in *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, 2016, pp. 1–7.

[14] D. Hovy, "The enemy in your own camp: How well can we detect statistically-generated fake reviews–an adversarial study," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 351–356.

[15] E. I. Elmurngi and A. Gherbi, "Unfair reviews detection on amazon reviews using sentiment analysis with supervised learning techniques.," *J. Comput. Sci.*, vol. 14, no. 5, pp. 714–726, 2018.

[16] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," *Journal of King Saud University-Engineering Sciences*, vol. 30, no. 4, pp. 330–338, 2018.

[17] C.-C. Wang, M.-Y. Day, C.-C. Chen, and J.-W. Liou, "Detecting spamming reviews using long short-term memory recurrent neural network framework," in *Proceedings of the 2nd International Conference on E-commerce, E-Business and E-Government*, 2018, pp. 16–20.

[18] R. Bhargava, A. Baoni, and Y. Sharma, "Composite sequential modeling for identifying fake reviews," *Journal of Intelligent Systems*, vol. 28, no. 3, pp. 409–422, 2019.

[19] C. G. Harris, "Comparing human computation, machine, and hybrid methods for detecting hotel review spam," in *Conference on e-Business, e-Services and e-Society*, Springer, 2019, pp. 75–86.

[20] M. R. P. Kashti and P. S. Prasad, "Enhancing nlp techniques for fake review detection," *Int. Res. J. Eng. Technol.(IRJET)*, vol. 6, pp. 241–245, 2019.

[21] G. Shahariar, S. Biswas, F. Omar, F. M. Shah, and S. B. Hassan, "Spam review detection using deep learning," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, 2019, pp. 0027–0033.

[22] Y. Fang, H. Wang, L. Zhao, F. Yu, and C. Wang, "Dynamic knowledge graph based fake-review detection," *Applied Intelligence*, vol. 50, no. 12, pp. 4281–4295, 2020.

[23] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "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*, vol. 90, pp. 523–537, 2020.

[24] S. Saumya and J. P. Singh, "Spam review detection using lstm autoencoder: An unsupervised approach," *Electronic Commerce Research*, pp. 1–21, 2020.

[25] H. Tang and H. Cao, "A review of research on detection of fake commodity reviews," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1651, 2020, p. 012 055.

[26] S. N. Alsubari, S. N. Deshmukh, M. H. Al-Adhaileh, F. W. Alsaade, and T. H. Aldhyani, "Development of integrated neural network model for identification of fake reviews in e-commerce using multidomain datasets," *Applied Bionics and Biomechanics*, vol. 2021, 2021.

[27] P. Devika, A. Veena, E. Srilakshmi, A. R. Reddy, and E. Praveen, "Detection of fake reviews using nlp &sentiment analysis," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2021, pp. 1534–1537.

[28] A. M. Elmogy, U. Tariq, M. Ammar, and A. Ibrahim, "Fake reviews detection using supervised machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[29] P. Gupta, S. Gandhi, and B. R. Chakravarthi, "Leveraging transfer learning techniques-bert, roberta, albert and distilbert for fake review detection," in *Forum for Information Retrieval Evaluation*, 2021, pp. 75–82.

[30] N. Jnoub, A. Brankovic, and W. Klas, "Fact-checking reasoning system for fake review detection using answer set programming," *Algorithms*, vol. 14, no. 7, p. 190, 2021.

[31] R. K. Kaliyar, A. Goswami, and P. Narang, "Fakebert: Fake news detection in social media with a bert-based deep learning approach," *Multimedia tools and applications*, vol. 80, no. 8, pp. 11 765–11 788, 2021.

[32] R. Mohawesh, S. Tran, R. Ollington, and S. Xu, "Analysis of concept drift in fake reviews detection," *Expert Systems with Applications*, vol. 169, p. 114 318, 2021.

[33] R. Mohawesh, S. Xu, S. N. Tran, *et al.*, "Fake reviews detection: A survey," *IEEE Access*, vol. 9, pp. 65 771–65 802, 2021.

[34] H. Paul and A. Nikolaev, "Fake review detection on online e-commerce platforms: A systematic literature review," *Data Mining and Knowledge Discovery*, vol. 35, no. 5, pp. 1830–1881, 2021.

[35] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Applied Soft Computing*, vol. 100, p. 106 983, 2021.

[36] M. B. Alvi, M. Alvi, R. A. Shah, M. Munir, and A. Akhtar, "Machine learning-based fake reviews detection with amalgamated features extraction method," *Sukkur IBA Journal of Emerging Technologies*, vol. 5, no. 2, pp. 10–17, 2022.

[37] H. Tufail, M. U. Ashraf, K. Alsubhi, and H. M. Aljahdali, "The effect of fake reviews on e-commerce during and after covid-19 pandemic: Skl-based fake reviews detection," *IEEE Access*, vol. 10, pp. 25 555–25 564, 2022.

[5] [19] [9] [7] [7] [12] [10] [8] [4] [2] [11] [10] [3] [6] [1] [20] [27] [23] [28] [14] [16] [30] [37] [15] [24] [21] [13] [35] [31] [36] [29] [33] [28] [34] [22] [26] [17] [18] [32] [25]