A NOVEL FAKE REVIEW AUTHENTICATION MODEL FOR ONLINE SOCIAL NETWORK USING CUSTOMER REVIEWS

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Applications

By

Vikas

Registration Number: 41800070

Supervised By

Dr. Isha Batra(UID:17451)

Department of Computer Science and Engineering

Lovely Professional University

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2024

DECLARATION

I, hereby declared that the presented work in the thesis entitled **"A NOVEL FAKE REVIEW AUTHENTICATION MODEL FOR ONLINE SOCIAL NETWORK USING CUSTOMER REVIEWS"** in fulfilment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision Dr. Isha Batra, working as Professor , in the Department of Computer Applications, Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Vikas

Registration No.: 41800070

Department/School: Computer Applications

Lovely Professional University,

Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled **"**A NOVEL FAKE REVIEW AUTHENTICATION MODEL FOR ONLINE SOCIAL NETWORK USING CUSTOMER REVIEWS**"** submitted in fulfilment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the Department of Computer Applications , is a research work carried out by Vikas (41800070), is bonafide record of his original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor) **(Signature of Co-Supervisor)**

Name of supervisor: Dr. Isha Batra Name of Co-Supervisor:

Designation: Professor Designation:

Department/school: Department of Computer Science

and Engineering Department/school:

University: Lovely Professional University University:

Abstract

Businesses can target their most loyal clients as the social network grows by analysing data and researching the most influential individuals. Customers are aided in their decision-making process regarding what to buy, where to obtain it, and which choice to make by reading ratings and reviews posted on various websites. Although locating authentic assessments online is becoming more common, more is needed to guarantee they are legitimate.User reviews are a crucial indicator of the worth of a product in the world of e-marketplaces. A significant number of businesses and suppliers of products employ spammers to publish deceptive reviews that attract new clients. There are three different kinds of fake reviews: reviews that are not true, reviews that are for a brand, and reviews that are not reviewed. Each of the three types misleads new clients. The vast majority of evaluations and ratings are made up, which leads to confusion among sincere customers and has a detrimental impact on their purchasing decisions. The reliability of online customer ratings and comments (OCRs) is becoming an increasingly pressing issue for society. However, the prevalence of deceptive feedback on online platforms substantially impacts the validity of e-commerce and the trust that customers place in online businesses. Studies already conducted to identify fake reviewers concentrate on extracting novel behavioural and linguistic characteristics. These features demand enormous human effort and experience, which places significant pressure on platforms. As an outcome, we present an innovative end-to-end architecture that can identify fraudulent reviewers based on their actions and the information they provide in their reviews. It comprises two essential elements: (1) a behaviour-sensitive sentiments feature extractor that learns the underlying patterns of reviewing behaviour; (2) a context-aware attention mechanism that extracts valuable features from online reviews. We comprehensively evaluate each suggested module and the complete framework by comparing it to state-of-the-art benchmarks using a real-world dataset obtained from https://osf.io/tyue9/. In this research , we discuss the technical elements of the most recent breakthroughs in neural networks-based word embeddings, highlight the primary merits and cons that these developments address, and look at how these technologies may progress. This study has two primary goals: the first is to identify fraudulent reviews accurately; the second is to assess the polarity of feelings

expressed in reviews; doing so will assist in enhancing the process of recruiting new customers. Both of these goals are interrelated. Buyers can come up with better buying choices and contribute to developing an inviting atmosphere over online businesses when fake reviews are revealed. In recent years, pre-trained language models have significantly improved the overall performance of natural language processing activities. Traditional word vector approaches such as Word2Vec cannot tackle the problem of a word having several meanings; however, these models can generate distinct representation vectors for each term in different circumstances; as a result, they are more effective at collecting the information contextual to the text. In addition, we consider that reviews typically include a wealth of expressions of opinion and sentiment, in contrast to most pre-trained language models, including RoBERTa, which needs to consider sentiment knowledge during the pre-training stage. Based on these variables, we offer a new model for identifying fake reviews named FRARBiLSTM(Fake Reviews-AFINN RoBERTa using Bidirectional LSTM). This model is built on a pre-trained language model and uses Bidirectional LSTM. Identifying Fake Reviews is made possible by the FRARBiLSTM model, which takes into account AFINN, RoBERTa, and BiLSTM. AFINN is a pre-trained language model that is based on sentiment knowledge augmentation. The research paper initially demonstrates employing Hybrid/Ensemble-based approaches called CatBoost+RoBERTa+AFINN that contrasted with preexisting traditional myths. Second, we presented a model that we called the FRARBiLSTM.Both of these methods perform far better than more conventional classifiers, with an accuracy of 92 and 97.31 percent, respectively, when identifying fraudulent reviews. So our model FRARBiLSTM, when used in conjunction with Ensemble Learning, achieves results that are superior to those achieved by even the most advanced word embedding algorithms and other deep learning techniques.

Acknowledgement

First of all, to our research supervisor Dr. Isha Batra, thank you for your extreme attention and invaluable guidance and feedback throughout this research work to complete. This thesis would not be in this phase without your commitment.

To everyone at Lovely Professional University, thank you for being nothing but supportive and helpful during the research period. My special thanks to Dr. Arun Malik for her kind guidance in completion of the research writing task, special guidance in the research papers publications with technical ideas in implementation part of this thesis. I would also like to thank the IPR cell of the Lovely Processional University to provide necessary tools for the practical part of the thesis and all assistance in filing the patents directly outcome of this Ph.D. research work entitled "A NOVEL FAKE REVIEW AUTHENTICATION MODEL FOR ONLINE SOCIAL NETWORK USING CUSTOMER REVIEWS" submitted in fulfilment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the Computer Application.

Last but not least, to our friends, family, colleagues, seniors and loved ones for being there for us during this more then five years of the time duration, without your support, it could not be possible to achieve this task.

(Signature of Research Scholar)

Name of Scholar : Vikas

Designation : Research Scholar

Department/school : Department of Computer Applications

University : Lovely Professional University

Date: April 2024

Table of Contents

Table of Contents

Table of Contents

List of Tables

List of Figures

List of Figures

List of Figures

List of Appendices

Chapter 1

Introduction

In the modern era of social networking and e-commerce websites, individuals are encouraged to submit their feedback and thoughts as reviews for any product, issue, or organization. Due to the enormous influence of reviews on consumers, spammers would use phoney reviews to advertise their goods or organisations while diminishing the image of their competitors' businesses. Online shopping has simplified the time-consuming task of visiting offline shops and picking items from a constrained inventory. Today's society makes it possible to purchase virtually anything online, from the most fundamental necessities to more expensive electronic appliances. Vendors abuse these big online platforms more frequently to increase the number of products sold by publishing bogus reviews. According to research on consumer interaction, approximately 82% of customers check online reviews before purchasing a product online. So, these reviews are essential for them to determine whether or not the product is trustworthy and whether or not it meets their needs.

Internet is readily available in urban and rural locations. This has also taken commercial matters to the Internet, which benefits both the consumer and the firm. People can rapidly share their product reviews, opinions, and experiences in discussion forums, blogs, and other social platforms. This type of content is known as user-generated content. No moderation exists because everyone is free to write whatever they wish. Because readers learn about someone else's genuine experiences with a product, reviews like these can sway consumers' purchasing decisions. The purchasers now consider these reviews an essential component of the shopping process, regardless of whether they purchase a new or current product. In today's day and age, product and service evaluations written by customers have become increasingly common. There are also lots of websites and applications specifically for customers to post reviews about their experiences with the products and services. The reviews express one’s personal experience with specific products and services. Many people make decisions based solely on the opinions of others. Therefore, more and more people post their views online.

The weight of reviews posted on the web is growing steadily. Reviews have the potential to sway consumers' final judgments. Customers can share their opinions, whether satisfied or dissatisfied with a product's positive or negative. Reviews like this are helpful because they give consumers insight into the experiences of other people who have already tried the product and can attest to its quality. Customers typically browse customer reviews before making a purchase decision. Buying probability increases significantly if the reviews are mostly good. Consumers often choose an alternative product when they see a preponderance of negative feedback.

While it is true that online reviews can be informative, both the buyer and the seller should avoid placing too much stock in them. Almost everyone who shops online first checks out consumer reviews. This helps when basing a purchase choice on online evaluations because some may be intentionally misleading for financial gain. Companies actively persuade consumers to write positive product evaluations to boost product sales.

Many aspects of private and professional lives have migrated to various online platforms throughout the last few decades. Similarly, burglars increasingly use the internet for new techniques to trick their victims. This includes deception committed during online purchases, which may result in (severe) implications for the buyer's body, feelings, and mind. In 2022, the German government agency for the Security of Information conducted a poll with a representative sample of the country's population and discovered that 8% of those polled had been the target of fraudulent activities concerning online purchases. The effects of this fraud ranged from monetary losses to losses in consumers' faith in various brands and purchasing platforms, as well as expenses in terms of wasted time[1]. Regrettably, this research cannot differentiate between the many types of online shopping fraud and does not define how these expenses could be allocated to the various categories of cybercrimes [1][2]. However, given the significant costs that are known to be involved with shopping fraud in general, the expenses connected to fraudulent online reviews are anticipated to be substantial [1][2].

1.1 OSN

OSN is a collection of nodes (individuals, participants, businesses, countries, territories, etc.) connected through links (relationships, relationships, boundaries, hyperlinks, etc.). Nodes can be individuals, actors, organisations, nations, states, etc. Web applications known as open source networks (OSNs) enable users to engage with one another, work together, and share material. How people think, express themselves, and interact socially with the outside world have all been impacted by OSNs. Nowadays, people carry out social and professional activities using various online campaigns and Research Gate. OSNs are essential to researchers and other fields because their structure is comparable to real-life communities and stores such user content. These fields include marketing, sociology, and politics, among others. Sociologists analyze OSNs as a tool for understanding human behaviour, and marketing firms use this knowledge to develop viral marketing tactics and reach new customers.[4][5][6]

Being a reliable communication route, it has become a widely used instrument for global events, which has improved the flow of information to share things. OSNs facilitate users' connections with professional groups in addition to helping them maintain their connections with friends and family. People sign up for OSNs and establish connections with other network users so they can stay in touch with one another and discuss topics that interest them in common. The ability for users to develop their profiles is one of the most fundamental and standard characteristics shared by most OSNs. Users can upload information to their profiles, including text, images, videos, and other media. They may be viewable by all members of the network or by members of the network that the user chooses. OSNs also make it easier for users to connect with other users and network members. When there is a connection between two members, it is assumed that those members are friends (or neighbours). Users near one another typically share interests. People can, for instance, strengthen their social relationships with their families and friends with the assistance of social networks. Twitter, Facebook, and MySpace are examples of online social networks (OSNs). Users are provided with a platform through which they may connect and share information with their connections via Facebook, which is the most popular online social network. Twitter enables its users to broadcast their

views, opinions, and suggestions while allowing them to receive updates from other linked users.

1.1.1 Merits And Demerits Of OSN

The most popular websites on the globe include Facebook, YouTube, and Twitter. Users have access to various tools that allow them to read and share material with their friends and contacts and search for future users who may have interests comparable to their own. If they are utilized appropriately, social networking sites offer many benefits to their users. On the other hand, if you do not use social media networks strategically, they can have a lot of negative consequences for your life.

Merits	
Connectivity	Individuals can easily stay in touch with their close ones. Social networking allows community linking who share similar interests.
Real-time access information	Fast and easy access to information from all over the world is provided by socialmedia.
Education	The education domain has had a positive effect on social media world wide. More easy for individuals to increase their domain knowledge via connections through social media sites to experts and professionals.
Great marketing channel for business	Social media marketing is the term used on social networking sites or on websites such as Facebook, Instagram, Tiktok,Twitter, YouTube etc. social media marketing.
Instant information	A huge advantage of social media is the opportunity to allow knowledge reach vast numbers of people.

Table 1.1 : OSN Merits

DEMERITS	
Privacy and Security Issues	Information on personal and security information can be easily hacked and shared on the Internet. Another issue that can cause a person financial loss by hacking their own personal
Cyberbullying	It has become very convenient for anyone to use the Internet because anyone can create a fake account and do something without being tracked down.
Frauds and Scams	Theft and scams occur on social networking sites, computer viruses, fraud and identity theft. The platforms are used to spread their beliefs and attract new members by exterminated hate groups.

Table 1.2 : OSN DeMerits

1.2 E-Commerce

The trend of doing more and more buying online is evolving daily. Websites dedicated to electronic commerce (e-commerce) opened a new market for buying and selling things online. Users can purchase goods (such as headphones, laptops, and so on) or use any service (such as hotel reservations, airline ticket bookings, etc.) through e-commerce websites. Many blog users, including users of e-commerce sites, contain user experiences of items or services in their blogs. New customers and product providers can benefit from the knowledge provided in the reviews, opinions, and suggestions that are placed online. People frequently consult their friends, family, and coworkers for recommendations before making any purchase. People can determine whether or not the product is worth purchasing with the help of these tips. Our choice of whether or not to buy a product is determined only by the feedback we receive from those who have personal experience with the product in question.

The practice of purchasing most of one's products online is evolving daily. Websites facilitating electronic commerce (e-commerce) online opened up a new market for buying and selling things. Users can purchase the goods (such as headphones, laptops, and so on) or use any service by using e-commerce websites (i.e., hotel reservations, airline ticket booking, etc.). After using any goods or services, users will frequently make ideas, thoughts, reviews, or comments on sites to put their experiences. Many blog users, including users of e-commerce sites, contain user experiences of items or services in their blogs[6]. New customers and product providers can benefit from the knowledge provided in the reviews, opinions, and suggestions that are placed online. People frequently consult their friends, family, and coworkers for recommendations before making any purchase. People can determine whether or not the product is worth purchasing with the help of these tips. Our choice of whether or not to buy a product is determined only by the feedback we receive from those who have personal experience with the product in question.

Customers can publish their ratings and comments on various websites today, thanks to the proliferation of the internet[6]. These reviews benefit not only the companies being reviewed but also potential customers for help in buying decisions. A noticeable growth in customer reviews has been observed over the past several years. The evaluations composed by previous purchasers impact the choices made by

potential buyers. Consumers reach a purchasing decision after perusing the assessments on social media and having already formed an opinion regarding the product. Therefore, customer evaluations offer invaluable assistance to individuals.

1.2.1 Conventional E-Commerce Models

Throughout its history, e-commerce has given rise to several distinct model implementations. The distinction between the buyer and the retailer is the most important criterion for classification. Traditionally, four primary categories can be utilized to categorise different business models used in online commerce.

B2B- In the B2B business model, one company offers a product or service to another company as part of its revenue generation strategy. The buyer and the final consumer are typically the same person; nonetheless, there are situations in which the buyer delivers directly to the customer. In general, the selling lifecycle for business-to-business (also known as B2B) transactions is substantially longer than that of consumer-to-consumer (C2C) transactions; nevertheless, the standard order amount is significantly higher, while there are frequently more repeat orders.

B2C- Businesses that market directly to consumers market to those using their products. The business-to-customer (B2C) model is by far the most prevalent; consequently, within this overarching framework, many distinctive ways exist. Customers can discuss their needs and concerns directly with the computer system that the company uses. In its most basic form, it is a form of online commerce that uses the Internet as a platform for customers to purchase any product they choose. They could also utilise online banking and booking services to pay their insurance premiums and utility bills online. The term "business to consumer" refers to the buying and selling of tangible goods and services. The selection process involved in purchasing a business-to-consumer (B2C) product is much quicker than the decision-making process involved in the B2B.

C2C- In a business model known as customer-to-customer (C2C), often known as an online marketplace, customers are connected by exchanging goods and services through fees for displaying their wares or using the platform. During the initial stages of the World Wide Web, internet-based companies such as Craigslist and Amazon became amongst the first companies to introduce the concept. C2C enterprises profit

from a marketplace powered by motivated consumers and vendors, although they have a significant challenge regarding their oversight of quality and keeping track of technology.

C2B- Conversely, companies that provide services to other businesses are the company's customers. Regarding C2B, the decision is made to take care of the customer or user first, then get down to business. Individuals can sell their wares or offices directly to other companies through C2B enterprises. In this version of the European Community, the platform would, in addition to providing opportunities for businesses, make it feasible for customers to complete the work they want. Affiliate marketing programs are another example of a C2B relationship. Recent pioneers in the field of innovation have made inventive use of this paradigm to connect companies with social media influencers to sell the companies' products.

The following are a few unique characteristics of e-commerce that make it particularly admirable in its current state.

The exploitation of New Business- E-commerce, in its most general sense, emphasises new opportunities and frequently uses language such as "develop economic ethics" and " perform more while doing less." Thanks to EC, customers have a greater chance to have input regarding the products developed, the processes by which those products are made, and how solutions are delivered. Customers had a greater demand for awareness of what was happening within the organization before the advent of EC. Nevertheless, it is a quicker and more open process, and clients have greater control over it.

Enhancement of Business Transactions- One of EC's primary missions is to improve business transactions across various platforms.

Production Efficacy- This, in turn, leads to improved results, such as enhanced efficiency, enhanced loyalty among customers, or optimal decision-making for enterprises. Consequently, this leads to improved outcomes.

Superior Financial Performance- Electronic trading can generate excellent economic performance (lower costs) and quicker business (high velocity, fast, and

immediate form interactions). Both of these goals could be accomplished by lowering transaction costs.

Execution of Data- This characteristic makes it feasible to execute off-data-loaded interactions that comprise multiple parties and take place via linked networks. These types of networks could include a mix of plain old telephone service (POTS) (also known as POTS), satellite television, rental telephones, or wireless technology. IB transactions give rise to the newest trends and sometimes even new business forms.

Transaction Incorporation- The processing of transactions is a common component of electronic commerce. Transactions are used to coordinate, monitor, carry out, and document other transactions. In addition, it allows clients to do electronic shopping and exchange money.

The Maximisation of Profits- Companies implement strategies to maximise profits by minimising price monitoring or increasing product sales. EC can increase revenue by generating modern markets for older items, building the most up-to-date IB materials, or constructing the most up-to-date networks of service distribution to provide consumers with the best possible service and engage with them. The capability of EC to manage transactions also enables businesses to reduce the time spent monitoring prices.

Friction Reduction- The primary objective of research on electronic commerce (EC) and the accompanying procedures is to reduce the dissipation during online transactions. In economics, friction is also referred to as transaction costs. This can be accomplished through marketplace structures and efficient combinations of technical programs, both essential for an efficient operation. If you encourage interactions between buyers, mediators, and sellers, conducting business online without encountering any long-term friction will be feasible.

Making It Easier to Form Networks:-This trend also impacts business connections because it is easier to build networks thanks to e-commerce. It encourages the formation of networks in which small businesses that are nimble and able to adapt to shifting conditions can more successfully rely on other partners, supply components, and distribution organisations to service to alter client demand. These networks can help small businesses compete more effectively in the global marketplace. Smaller

enterprises may establish these networks that can respond effectively to shifting circumstances. The management of the network chain that connects customers, employees, vendors, distributors, and even competitors needs to be one of the goals of the end-to-end relationship management approach.

Making a New Organisational Paradigm Possible- This makes a new organisational paradigm very different from the one that existed in the past. It is an establishment that provides control for an IB organization. Improving management roles, working group structures, and the flow of knowledge and collaboration are some emerging models of the techno-organisational framework.

1.3 Importance of User Reviews in OSN and E-commerce

Under the current conditions, consumers are forced to depend more on making decisions to acquire products through e-commerce websites or traditional retail establishments. Reviews are manipulated to produce either favourable or wrong impressions to influence the success or failure of a product's sales. This is happening since reviews significantly impact the success or failure of a product's sales. Some names for reviews have been tampered with, including fake reviews, fraudulent reviews, opinion spam, and untruthful reviews.

Dishonest opinion spam has developed into a problem in today's digital environment, posing a threat to customers and enterprises [3][6]. Locating these fake reviews is an essential and challenging task that needs to be accomplished. Significant shifts have occurred in how multinational companies slander the products of other enterprises operating in the same area. An investigation carried out not too long ago by Samsung's Open Tide had paid persons to make inadequate internet evaluations against HTC, which encouraged Samsung cell phones. The individuals who authored the assessments emphasised what they saw to be deficiencies in HTC's goods while downplaying any unfavourable aspects of Samsung's offerings.

Amazon.com, the biggest e-commerce site in the world, admitted not long ago that it had published fake customer reviews on its website. Following this, the company initiated legal action against three other websites, accusing them of engaging in fraudulent rating practices and demanding that they cease doing so. Fakespot.com has established itself as the industry leader in detecting false reviews of products

advertised for sale on Amazon.com and other e-commerce sites that are a part of the Amazon subsidiary network. Reviews and ratings can impact customers' purchasing decisions directly. They contribute significantly to the prosperity of companies. Negative reviews can damage a company's reputation and lead to financial losses. Positive evaluations, on the other hand, can lead to improvements in economic standing.

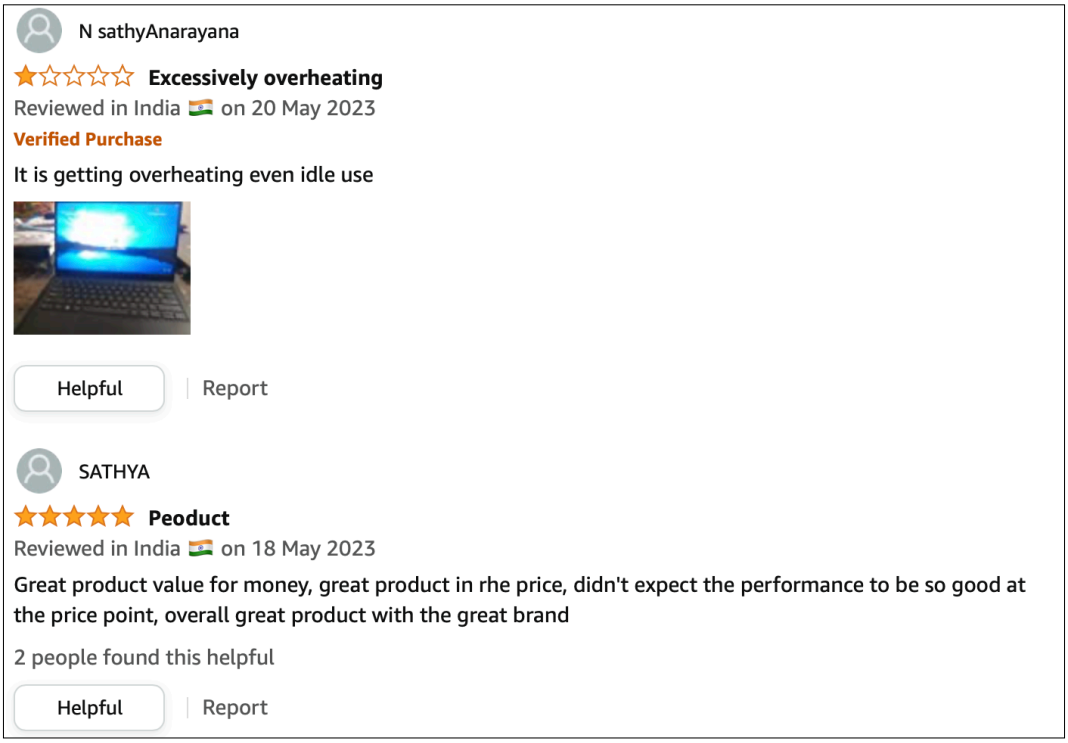


Figure 1.1: Reviews on Amazon

The number of people purchasing through online marketplaces like Amazon and eBay is growing daily. Customers who make purchases online frequently write reviews and share their thoughts on the products they have bought and utilized. User opinions are pieces of material generated by users on e-commerce websites to represent the experiences that consumers have had with any given service or product. Customer feedback can be analysed from both the customer's and the company's point of view. These reviews can impact the buying choices of potential new consumers or users for a particular product. Reviews from existing customers significantly influence the choices made by prospective new customers.

In user reviews, the product's positive and negative aspects are discussed according to the reviewer's own experience, which assists other users in helping to buy the item. Users frequently check large e-commerce sites with user experiences about

products before purchasing online. The quantity and quality of the user experience influence the amount of user traffic on a site.

Reviews represent the suggestions, opinions, or experiences of someone about anything now available for purchase. Users give product providers, sellers, producers, and new customers feedback in the form of reviews they post on e-commerce websites that are accessible online. These reviews are intended to offer suggestions or share experiences. Each business that sells a product can either make or lose money depending on the polarity of the evaluations customers give that business.

On the other hand, new buyers are influenced by reviews when making decisions on which products to purchase. The consequences of reviews have a variety of effects on businesses and individuals. Keeping this viewpoint in mind, many companies and product providers use agents to generate fake opinions to grow their businesses and enhance their standing in the market. Consequently, users need better choices regarding the products they want to purchase.

The most significant benefit of favourable customer reviews is that they increase potential customers' confidence in the company. When a customer has high confidence or trust in your company, he is more inclined to suggest it to others in his circle [4][5]. A significant amount of money is allocated to a company's advertising efforts. People put a lot of stock in the reliability of online reviews. Hence, receiving favourable ratings is equivalent to receiving free advertising.

1.4 Fake Review

Throughout the past several years, there has been consistent growth in the number of online shopping stores. The identification of fake reviews has garnered attention due to the enormous expansion seen by these businesses[5][7]. Customers are being intentionally misled by fake reviews, which undermines the honesty and authenticity of venues where people shop online. Opinion spamming refers to submitting deceptive reviews and is immoral behaviour. The purpose of opinion spamming is to confuse the reviewers with false information. Users that engage in spamming activities are referred to as "spammers." The job of spammers is to create a fictitious reputation (either positive or negative) for a company by posting fake reviews about

the company. A review has "positive polarity" if it contains a comment promoting the product. In addition, evaluations that contain extremely unfavourable statements about the product are said to have "negative polarity."

As reviews impact potential customers' decisions, online reviews have become very important for businesses in attracting customers. Sometimes, merchants exploit people's trust in online reviews, propagating misleading positive reviews to boost their sales. In addition, some merchants even spread negative rumours about competitors to increase their market occupancy. Therefore, there are two types of fake online reviews: fake positive and fake Negative. A review that contains lovely comments but does not honestly represent how good or poor a product is is considered a fake-positive review. A review that contains negative comments but does not accurately assess how good or poor a product is is deemed to have a fake negative review.

Because these ratings impact customers' purchasing decisions, some individuals post false reviews on e-commerce websites to boost sales of their products. These individuals are generally called opinion spammers, and their behaviour is called opinion spamming. Every day, more and more examples of fraudulent reviews are posted online. Some retailers are capitalising on the opportunity to expand their businesses more rapidly by paying opinion spammers to create fake reviews of their products. Many websites pay people to post fake reviews on many other websites. As a result, identifying these fake reviews has developed into a significant challenge in preserving the confidence between the client and the buyer.

Customers who create fake ratings and reviews can potentially destroy a business's reputation. It can affect how others perceive or purchase a product or service. Identifying fake or fraudulent reviews has long been accomplished through the use of conventional methods of data analysis. The earliest data analysis methods primarily focused on extracting quantitative and statistical data properties. To outperform the capabilities of a traditional system, a data analysis system needs to be pre-loaded with a significant amount of background data and can reason using that data. This is necessary for the system to be considered advanced. Researchers have focused on machine learning and artificial intelligence subfields to achieve this objective. Review authenticity can be established using supervised or unsupervised learning

methods. These approaches generate user profiles primarily through cookies to search for reviewer profiles, review data, and activity on the reviewer's part while using the internet. Whether we use a supervised or unsupervised technique, we all get a rough estimate of how likely fraud has been performed.

There is no single statistical study that can definitively determine whether or not a specific review is fraudulent. This can only point to one conclusion: this review is more likely to contain fraudulent information. Researchers and e-commerce websites face a unique challenge when it comes to the detection and filtering of genuine customer feedback.

After looking at reviews on several different e-commerce websites, researchers concluded that over 10 percent of the ratings on websites selling goods online are completely fabricated [5][8]. After analysing evaluations found on various e-commerce platforms, the researchers concluded. In addition, it was found that more than 75% of all false reviews have a positive polarity. Researchers are focusing their emphasis on finding a solution to the issue of recognising bogus reviews because the necessity to do so is growing. Not only can fake reviews confuse new customers when deciding which products to purchase, but they also hurt the business of companies that make high-quality goods. Users will only go to a specific e-commerce website if it has a lot of fraudulent and misleading reviews. It has been determined that spotting false reviews will, at the same time, prevent three separate losses.

Consumers can be duped into buying a product if they read reviews made to mislead them. The term "fake review" refers to this particular kind of review. In many cases, the people who write these reviews have yet to contact the goods or services being evaluated directly. In today's world, most firms maintain an online presence on social media to increase the number of macro conversions for their operations. Even customers have the advocacy function, which encourages them to seek out experience with the product through word of mouth. This assists in determining the Sentimental Polarity that is linked with the customer.

Additionally, Loyal Reviews are a significant factor in determining businesses' Online Reputation Management (ORM) status. Customers' thoughts and attitudes are

increasingly influenced by the online reviews they read before purchasing products or services. Every review is dishonest and intentionally created by an individual or a team to praise or criticise a particular product. Nowadays, when most individuals prefer to read product reviews on online social networking sites before purchasing, how businesses operate is significantly impacted. The reliability of a review can be determined through a method based on machine learning by examining the degree to which two words or passages are alike. The presence of online reviews significantly impacts the procedure that businesses use to acquire new customers. During our investigation, we came up with a few machine learning algorithms that split the review to determine both the grouping of the review and the parameterised element of the review. These algorithms were developed as a result of our work. This part of the review contributes to the determination of the next part of the work to determine whether the review is false.

Unjustifiably good reviews can promote a product, while unjustifiably negative reviews can undermine the reputation or sales of competing products and companies. Either way, the evaluation aims to promote the product in question. These results are designed to encourage the evaluated product [6][9].

Review spam can be broken down into three distinct types: corporate reviews, non-reviews, and deceptive reviews. A particular target product's reputation can be falsely promoted or demoted through the use of untruthful evaluations. There is a possibility that statements about various aspects of the product could be included in the text of fake reviews. Posting negative brand reviews criticises the company that provides, manufactures, or distributes the goods. The information in posts that do not review is irrelevant to the product with a review written on it. The content of items not reviews may consist of question-and-answer sections or advertisements.

There are primarily three categories of spam reviews-

Reviews made to praise or criticise a product- Most reviewers have never used the services or items they are critiquing.

Reviews that focus solely on brands- Instead of concentrating on one specific product or service, in this kind of review, the author talks about the company that makes the thing being evaluated, often known as the brand. For instance, a

smartphone review by Samsung might include something like, "I hate Samsung phones, and I have never bought one of their devices."

Non-reviews- These evaluations are nothing more than advertisements and unrelated content, and they do not offer any perspective that is even remotely reasonable.

Two categories of people write fake reviews-

Professional fake reviewers- Many reviews are often produced when working with a professional fake reviewer. They might do this as freelancers or work for a company that writes fake evaluations. They are compensated monetarily for the work that they perform. Since they write many reviews, various data mining algorithms can quickly discover their linguistic and behavioural tendencies. This is because they write reviews. However, they may have already caused considerable destruction when they were found. When detected posting fake reviews, they delete their old account, open a new one, and immediately begin posting fake reviews under the new one. This makes the problem even worse.

Nonprofessional fake reviewers- These individuals will create fake evaluations to benefit themselves and their friends and family members. Most of the time, they do not receive compensation for their labour. Because they only post a few fake reviews, it isn't easy to detect this account's linguistic and behavioural characteristics.

The term **"sentimental polarity"** relates to the notion that a product is either considered high quality or low quality based on the generally employed polarity parameters of evaluation, including positive, neutral, and negative reviews. People can post their activities, disseminate information, evaluate products, share images online, offer product reviews, express their opinions and thoughts, and share any text, audio, or video content on the OSN websites. People can also give product reviews. The majority of people today communicate their feelings and thoughts on social media. For example, if individuals do not like a particular firm's service or product, they immediately share their comments about the product or service. Businesses now take this kind of behaviour very seriously on social media platforms because of the potential damage it could cause to their reputation on the internet. Customers are pleased with the business firm's Page(s), and after using the company's product or service, customers provide reviews describing their experiences with the product or

service. As these reviews are provided in the comment section, other consumers see them before purchasing the same product or service from the same firm. As a result, these reviews have the potential to influence the purchasing decisions of other customers who are planning to buy the product.

Document Level		
Text		**Sentiment**
Few days ago, I purchased a laptop. It's such a solid hardware. The processing speed is also so fast. I simply love this product.		Positive
Sentence Level		
Text		**Sentiment**
The laptop is awesome		Positive
Aspect level		
Text	**Aspect**	**Sentiment**
The Laptop processing speed is very fast, but its battery life is short.	Processing speed	Positive
	Battery life	Negative

Table 1.3: Reviews Sentiment Polarity

Because these are unstructured data written in any language, including English, only a little research has been done on them up until now. If these reviews are correctly analysed, then they can provide customers with the proper insights, and if a company is providing good service, then the company may earn huge profits from these reviews; however, if a company is providing lousy service, then customers can save their money by avoiding these companies. Analysis at the document, sentence, and aspect levels are the three levels that can be used to characterise the sentimental polarity of a statement. When analysing at the document level, attitudes are determined based on a single entity rather than several entities. Analysis at the sentence level can identify whether or not a phrase is under sentiment polarity. *Aspect level analysis* is an analysis that is based on the feature or aspect level of the entity. In this type of analysis, several feature aspects are selected from the text, and then the attitudes are evaluated.

In the current market environment, people rely more on reading reviews before purchasing things, whether online or in traditional retail outlets. Because of the

critical part that reviews play when it comes to assessing the effectiveness or lack thereof of a product in the market, reviews are increasingly being persuaded to reflect excellent or unfavourable ideas. This directly results from reviews' vital function in making this determination. Reviews that have been altered may also be called fake reviews that are not true. Each of these different names refers to the same thing: reviews that have been changed. Every term refers to the same thing, which is testimonials that have been fabricated. In the modern digital environment, spam containing inaccurate viewpoints has turned into an element that poses a risk to customers and businesses. Finding the difference between genuine and false reviews is a process that is both crucial and challenging. These dishonest reviewers are frequently compensated for their work in writing these reviews.

Consequently, it is tricky for an average client to determine which evaluations are legitimate and fake simply by reading through all the reviews. Significant progress has been achieved in regulating multinational corporations that smear the reputations of the items sold by rival businesses operating in the same market. A recent investigation proves that Samsung's unit, Open Tide, had paid people to write internet reviews that were friendly towards Samsung devices and adverse towards HTC handsets. These reviews were posted on the internet.

The authors of the reviews gave more weight to the issues they identified as being present in HTC's products than they did to any criticisms that could be levelled against Samsung's offerings. Amazon.com, the largest online retailer in the world, admitted not long ago that fake ratings had compromised its website. After that, the corporation took the next step and initiated legal action against three other websites, accusing them of publishing fraudulent reviews and requesting that they stop doing so.

Fakespot.com has solidified its position as the market leader in detecting fake reviews of products sold on Amazon.com and other e-commerce sites affiliated with Amazon.com. The company offers both a grade and a proportion of reviews that are determined to be fraudulent. Customers' purchasing selections could be directly influenced by the evaluations and ratings left by previous customers. They constitute the fundamental elements that positively contribute to business success. A company's reputation and bottom line can take a substantial hit due to negative comments about

its products or services. Positive reviews, on the other hand, can lead to an increase in revenue. Customers who create fake feedback and ratings can destroy a company's credibility. Identifying fake or fraudulent reviews has long been accomplished through the use of conventional methods of data analysis. The earliest data analysis methods primarily focused on extracting quantitative and statistical data properties. For a data assessment system to have capabilities superior to a traditional design, it must have been pre-loaded with a significant amount of context data and can reason using it. This is important since we must give the impression that the system is technologically advanced.

Researchers have focused on machine learning and artificial intelligence subfields to achieve this objective. Using supervised and unsupervised learning approaches, a review can be determined to be either fake or real, depending on which technique is used. These approaches generate user profiles primarily through cookies to search for reviewer profiles, review data, and activity on the reviewer's part while using the internet. There is no single statistical study that can definitively determine whether or not a specific review is fraudulent. This can only point to one conclusion: this review is more likely to contain fraudulent information. Researchers and e-commerce websites face a unique challenge when it comes to the detection and filtering of genuine customer feedback.

1.4.1 Fake Review Parameters

In detecting fake review content, two primary features can be relied on: **behavioural and textual.**

Behavioural Feature of Fake Reviews

The quantitative relevance of a consumer's behaviour and feedback can be evaluated using behavioural features based on the individual's prior and current reviews. This can be done by analysing the user's review history. One needs to be familiar with symbols and formulas for statistical analysis. A concise summary of these characteristics is below.

F1-MAX REVIEWS

Most research revealed that 75% of fraudsters posted five or more reviews daily [7-9]. Nearly 90% of ordinary users never publish an additional rating every day.. As a result, the number of reviews each individual contributes can be used to differentiate between average reviewers and spam reviewers. The following formula can be used to get the maximum number of reviews.

$$F_1(a) = \frac{MaxRev(a)}{max(MaxRev)} \quad \text{.....(i)}$$

PERCENTAGE OF POSITIVE REVIEWS -F2

A spammer may have published fake reviews if the vast majority of the evaluations they have written regarding products are positive [10]. The following calculation can be used to sift out specific reviewers who seemed to support businesses and determine the percentage of reviewers who gave favourable feedback with a score of four or five stars.

$$F_2(a) = \frac{\sum_{x=1}^{|R_a|} |\{\star(r_x) \in \{4,5\}\}|}{|R_a|} \quad \text{..... (ii)}$$

AVERAGE REVIEW LENGTH -F3

Most of the research that has been conducted up to this point has demonstrated that spammers do not publish in-depth reviews about the services or goods they use, which may be utilized to identify spammers [7-9], [11-14]. Since the spammers' eventual goal is to generate fake reviews, the authors of the reviews themselves often take minimal effort on their part [15]. On the other side, ninety percent of reviewers who can be relied upon to provide accurate information provide more in-depth critiques, with a typical length of more than two hundred characters. The writers [15] concluded that the testimonies were fake and significantly shorter than X. The following formula, with X equal to 135, can be used to calculate the typical length of an evaluation.

$$F_3(a) = \begin{cases} 1, & len\,(r_a) < X \\ 0, & otherwise. \end{cases}$$ ..... (iii)

BUSRSTNESS (BST) -F4

To generate fast results, most spammers artificially inflate the ratings they receive. It is considered an odd action when many reviews are posted quickly, which may indicate that the user is trying to spam [16]. This approach is provided to analyze the total amount of the writer's evaluations produced by the person using the application in the preceding twenty-four hours before the current time. If the aggregate amount of reviews exceeds a certain threshold, the individual in question is likely a spammer. The value X = 28 was chosen for the threshold position after the research results on the experimental dataset were analyzed. The following method can be utilized to determine the burstiness characteristic:

$$F_4\,(a) = \begin{cases} 1, & \sum_{x=1}^{|Ra|} |\{r_x \in R_a\} \\ & \cap (t_x\ is\ in\ last\ 24\ hours| > X) \\ 0, & otherwise \end{cases}$$ ..... (iv)

REVIEWER DEVIATION -F5

A reviewer must rank the goods according to whether other consumers have evaluated them on average, and an objective reviewer will do so. On the other hand, if spammers try to downgrade or upgrade particular products, their ratings could be significantly different from the average ratings for such products.

It is common for spammers to give a product a high rating, which can make it easier for us to identify fake reviews [7], [9], [17]. Therefore, using reviewer divergence as a possible action that can be utilized for recognising fake reviewers [18] is something that can be done. A mathematical formula that helps to obtain the F5 of this characteristic.

$$F_5(a) = avg\frac{\lvert r_{ap} - \bar{r}_p \rvert}{4}$$	..... (v)

WEIGHTED RATING DEVIATION -F6

Early deviation is a term that refers to the activities of a fraudster who spams a review not too much longer after the good or service becomes publicly accessible to the general population. These spams will draw the focus to additional hackers, who would then capitalise on the viewpoints of successive spammers [15], [19]. It takes time for damaged products to recover from the damaging influence of these negative early reviews posted by another authorised reviewer. This recovery process can take up to several months. How the rating was initially allocated serves as a reflection of the weight that the rating carries. Calculating the rating's weight might take many forms, such as the one below.

$$F_6(a) = \frac{1}{(r_{ij})^{\grave{a}}}$$	..... (vi)

NEGATIVE REVIEWS RATIO -F7

It is necessary to ascertain the number of bad reviews a reviewer has contributed to a publication since the proportion of good reviews is essential. By determining a percentage of the reviewers' total negative input, the suggested strategy could screen out reviewers more likely to downgrade companies [15]. Because of this, the percentage of negative evaluations with a score of one or two was calculated. The following formula can be used to determine the less-than-favourable reviews ratio.

$$F_7(a) = \frac{\sum_{x=1}^{\lvert R_a \rvert} \lvert \{\star(r_x) \in \{1,2\}\} \rvert}{\lvert R_a \rvert}$$	..... (vii)

MAXIMUM CONTENT SIMILARITY-F8

A reviewer is highly likely to be a spammer [20], [21], and [22] if the review contains text that is identical yet pertains to different products. They will usually post similar reviews regarding numerous products to support their claims. This is done because they want to save time coming up with new, erroneous evaluations [23]. Because of this, it is vital to determine the content similarity between the author's reviews to recognise the author's spamming behaviour. The F8 characteristic is determined using the following formula:

$$F_8(a) = \max_{x}^{y} [cosine(r_i, r_x)] \, where \, r_i, r_x \in R_a, x < y \quad \text{..... (viii)}$$

REPEATED REVIEWS

When people publish multiple and repeating evaluations for the same product, this is a sign that they are engaging in abnormal behaviour [24]. The same person who wrote the reviews several times should not be classified as a false review because of problems with internet connectivity or operational defects.

F9 AND F10-BOTTOM-RANKED REVIEWS

Faithful consumers are more likely to evaluate a product or service after using it, which means they will spend more time doing so. This is in contrast to spammers, who rate quickly in an attempt to influence customers' choices [19]. The feature that displays the reviews with the bottom rank can be computed as follows.

$$F_{10}(a) = \frac{|\{r : r \in R_a \& F_{BRR}(r) = 1\}|}{|R_a|} \quad \text{..... (ix)}$$

TOP-RANKED REVIEWS RATIO -F11 AND F-12

If most reviews are high-ranking ratings, the reviewer's behaviour may be viewed as questionable [19]. Below is a demonstration of how to compute the top-rated review ratio feature.

$$F_{11}(a) = \frac{\lvert\{r : r \in R_a \& F_{TRR}(r) = 1\}\rvert}{\lvert R_a \rvert}$$	..... (x)

EXTREME RATING BEHAVIOUR -F13

Consumers can choose to improve or decrease the standard of the products based on their ratings from most excellent to least great. Similarly, spammers would give outcomes either positive or negative evaluations to promote or degrade them [9]. Extreme rating conduct includes providing one or five stars in a system that only allows for a maximum of five stars. The following formula can be applied mathematically to analyze the behaviour of extreme ratings.

$$F_{13}(a) = \begin{cases} 1, & \star(r_a) \in \{1,5\} \\ 0, & \star(r_a) \in \{2,3,4\} \end{cases}$$	..... (xi)

FIRST REVIEW RATIO -F14

Initial assessments of a company's services and products can significantly impact the sales of the company being reviewed and its competitors. Hackers also attempt to pose as early reviewers to gain more significant influence, which they use to deceive potential consumers [25]. The initial assessment ratio may be calculated using the example provided below.

$$F_{14}(a) = \frac{\sum_{x=1}^{\lvert R_a \rvert} \lvert\{r_x \in R_a \cap (r_x \; is \; a \; first \; review)\rvert}{\lvert R_a \rvert}$$	..... (xii)

Textual Features of Fake Reviews

Text features comprise the semantic, grammatical, and linguistic characteristics of the review, in addition to the metadata characteristics. These characteristics can be used to help identify bogus reviews. Numerous aspects fall under this category, each of which will be broken down further in the following discussion:

META-DATA

These parts consist of authentic reviews and information about reviews, such as comment ID, suggestions, article length of sentence, score, information, user ID, and retailer ID [10], [22]. It has been established that making use of meta-information features is effective for detecting fraudulent reviews. Using meta-data information allows for the detection of unusual or aberrant reviews. When a reviewer is discovered to produce fake reviews, it is simple to classify all reviews connected to that reviewer as counterfeit. Unfortunately, these features might not be found in every data source, severely limiting their usefulness for identifying fake reviews. For instance, experts can identify a particular spammer based on the reviewer's profile. With the help of the reviewer's computer's IP address, it is possible to locate the precise position of the reviewer's System, the length of time individuals spend looking at the content, and the comments provided by the reviewer.

BAG OF WORD (BoW)

This pattern of behaviour, which can be perceived as suspicious, occurs when some users generate a large number of reviews of a single product, either favourable or unfavourable while using the same computer. When we compare the numerous brands of competing goods, among the activities that we do is take into account the ratings that reviewers have provided for specific products. These ratings can be found on the product's page. One reviewer has provided a significant number of remarks, the majority of which are positive, concerning a specific brand's items. The identical customer has also negatively reviewed other brands that sell goods in the same market.

PART OF SPEECH (POS)

One aspect of POS that can be utilised to one's benefit is the frequency at which each POS (Part of Speech) occurs in the source material [26], [27]. Previous investigations in computational language science indicated that various text formats exhibit varied word order and syntax patterns to a certain extent; the outcomes of this research supported this assumption that diverse text formats exhibit different word order and syntax patterns. In contrast to other features, such as BoW [29], [28], the POS feature could be more successful in identifying fake reviews as fraudulent.

LINGUISTIC INQUIRE AND WORD COUNT (LIWC)

This is an established piece program for analysing texts that may be applied to a text to investigate the linguistic characteristics of that text from various perspectives [104–106]. When detecting false reviews, this function is not nearly as effective as others, such as the unigram, bigram, and trigram checks [29], [28]. Enhancing model accuracy could be accomplished by utilising LIWC to add n-gram features. To provide a more granular degree of insight, it counts the number of times the keywords are used. It sorts them into main psychological components, which are then divided into 4 primary groups: personality traits, language traits, speech characteristics, and behavioural characteristics. This framework was built for spoken language and was not extended to accommodate natural language; thus, you should refrain from utilising these capabilities. Using a different system is necessary if you wish to employ natural language.

STOLYMETRIC

These include syntactic traits, word-based characteristics, ranging features, and character-based properties [30]. Word-based assessments typically include information regarding the typical length of a word and the number of letters capitalised in that term. This indicates the types of words and characters utilized by the reviewer. It is possible to determine the reviewer's approach to writing based on syntactic aspects from the review, including the amount of punctuation tags used.

THE FEATURES OF SEMANTICS

This offers illustrative examples of the concepts that underlie words or the meanings that lie under their surface. Combining these aspects produces a semantically meaningful detection approach for bogus reviews. According to Li et al.'s [31] research on cross-domain performance, semantic features perform noticeably better than the rest of the models that support n-gram. After some time, Kim et al. [32] presented a method that utilised FrameNet-based semantic characteristics to demonstrate that the classification results had greatly improved. Nevertheless, these characteristics are not able to provide an accurate representation of the semantic link that exists between documents.

1.5 Analytics Models

Analytics modelling is essential for comprehending data, developing forecasts, and selecting appropriate courses of action for a firm. If no models are involved, drawing any relevant inferences from the data will be tough. Big data analytics' primary objective is to bolster company confidence to drive higher levels of productivity and profitability across an organisation's operations. *Business analytics* is a process that entails discovering key data patterns, analysing those trends, disseminating those analyses to others, and utilising resources to make it feasible for the entire organization to pose a query about any data in any environment using any technology. All of these steps must be completed to be successful. This article examines the current framework for prescriptive and predictive analytics and the primary methods for carrying out this framework.

In addition, it conducts research on various analytical methods, summarises previous research publications to identify ongoing inquiry concerns, and outlines potential future research paths. In analytics, information and mathematical computations are utilized to determine answers to issues, establish linkages, and foresee outcomes that cannot be predicted with certainty. This broad field of computer science generates new information and characterises fundamental data patterns. The success of this endeavour relies heavily on its foundations in applied mathematics, statistical modelling, and predictive modelling. Any business analytics program will attempt to maximise the efficiency of company operations by combining confidential company

information with data that is freely accessible to the public and additional data gathered from third parties.

The analysis of large amounts of data should not be considered a universally applicable approach. The more skilled a data scientist or data analyst is, the greater the possibility they will be able to track the kind of analytics firms could use to their advantage. As was covered in the accompanying discussion for Figure 1.2, there are four distinct analytical forms.

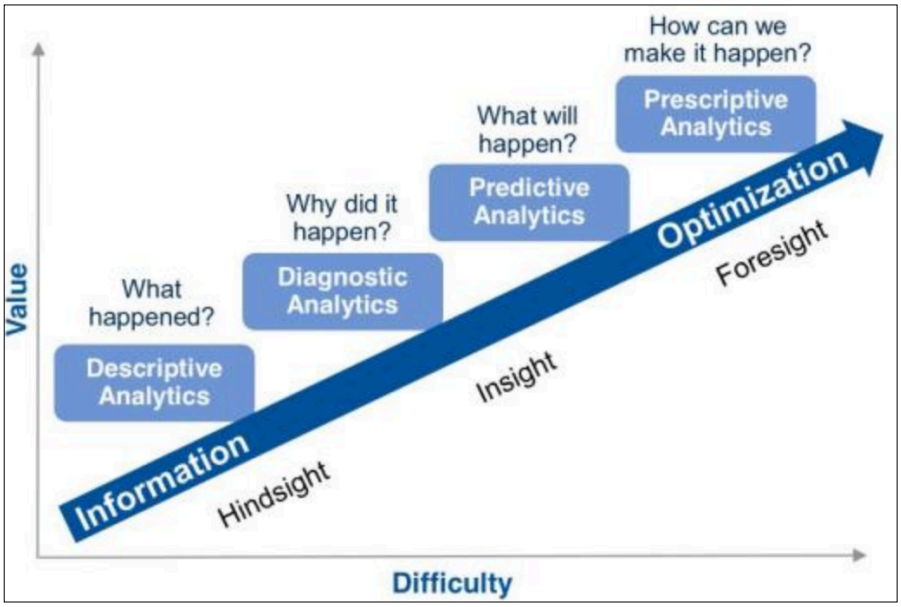


Figure 1.2: Analytical Models

DIAGNOSTIC ANALYTICS- This aims to answer the query "Why does this occur?" by analysing the information and materials. During the discovery phase, analytics will assist them in locating the data sources to explore the findings more accurately. Concentrating on a particular data component or widget constitutes "drilling down." You can quickly drill down to this level of detail using the Hisense BI platform. *Data mining* is a method that uses automation to extract useful information from vast amounts of unprocessed data.

DESCRIPTIVE ANALYTICS- The area of statistics known as descriptive analytics focuses on the collecting of raw data as well as summaries that are straightforward to comprehend. Descriptive research will frequently use past information extensively to provide the structure necessary for correctly evaluating both info and numbers. Descriptive analytics is looking at past data to understand how an organization has

changed. Therefore, by utilising previously collected data and benchmarks, decision-makers can acquire a comprehensive picture of the results and patterns on which they may focus their company plan. Because of this, they can make more well-informed decisions. The most fundamental application of descriptive analytics is producing reports constructed using information from online sources such as websites, blogs, and online datasets. As shown in Figure 1.3, several performance metrics for websites can be found in the descriptive analytics category.

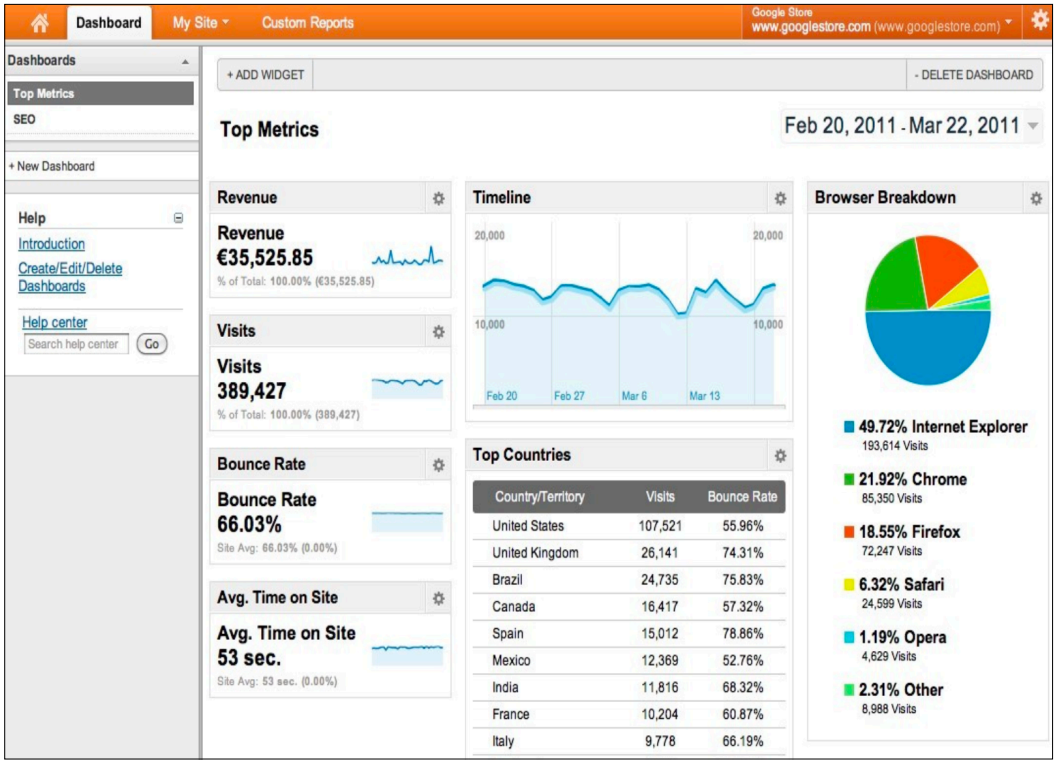


Figure 1.3: Key Performance Indicators

PREDICTIVE ANALYTICS- It is an area of predictive analytics that predicts unanticipated events in the future. These techniques are combined with others to generate reliable forecasts of the future based on existing information. It is a subfield of analysis that makes forecasts about events in the future that have yet to take place by employing a wide variety of methods, tools, and statistical approaches. It is referred to as predictive analysis. Artificial intelligence and data mining domains benefit significantly from applying machine learning and mathematical algorithms. The software does an exhaustive investigation into the data that is now accessible to find patterns of transactions and previous occurrences that are helpful to the organization in picking them. Transaction profiling, decision analytics and optimisation, and predictive modelling are the three critical components of predictive

analysis. When conducting predictive analytics, it is possible to discover challenges and opportunities using historical trends and transactional data. Data extraction from sets, spotting patterns and trends, and forecasting future events are likely results of predictive analysis, which uses data mining, mask learning, and other predictive technologies. Predictive studies are helpful when it comes to managing risk analysis projects. These studies conform to a method that outlines the project's description, collects and analyses data, evaluates gathered statistics, and creates predictive models.

PRESCRIPTIVE ANALYTICS- The massive amounts of data already accessible to businesses are the primary reason for this. Figuring out which approach to data analysis will be the most beneficial could be a difficult task. PA is one option that could help your company recognise data-driven strategic decisions and assist you with escaping the traps of typical data analytics approaches. If your company is interested in exploring these possibilities, go here. Instead of merely projecting what will happen to your firm, a prescriptive analysis modifies those aspects to achieve the most beautiful result and action plan. A predictive analysis, on the other hand, is limited to making projections about what will take place. Predictive analytics, or PA, is at the forefront of recent developments. The prescriptive analysis goes beyond merely forecasting possibilities in the predictive model and instead expressly proposes many proposed alternatives and possible ramifications to each behaviour. This is because the prescriptive analysis goes beyond simply forecasting options in the predictive model. This is because prescriptive analysis involves more than merely forecasting options within the predictive model. The goal of prescriptive analytics is to decide, depending on the information that is now available, the course of action that will be most productive. Instead of focusing on data monitoring, this analysis emphasises the observations that can be acted upon. It is generally agreed upon that prescriptive analytics is the next frontier in corporate analytics. It provides businesses with a path of action that is flexible, automated, and based on the passage of time to exploit future economic possibilities.

1.6 Natural Language Processing

A flood of data has been produced due to the growing digitisation of information and the volume of transactions. The world's data has doubled in size in a concise amount of time due to the constant acceleration of digital information speed. Natural Language Processing is a method that analytical models employ to comprehend, analyse, and process human language to organise hidden information. Construction of an NLP pipeline begins with sentence segments. The two primary techniques for analysing natural language are syntactic and semantic analysis. Word Sentences can be divided into tokens or individual words using a tokenizer. Stemming is used to normalise words to their root or fundamental form. The proper base form is created through lemmatization. Adjectives, verbs, adverbs, and other parts of speech are included. It concerns the grammatical and semantic function of a word in a sentence. Different words can be classified as other parts of speech depending on the situation. Chunking is a technique that groups smaller sentence fragments of information.

NLP techniques for extracting pattern-based information

A conceptual series of symbols or phrases with a specific regularity constitutes a pattern. A regular expression for tokens is a pattern. The tokens used in this pattern may be constant-precision string literals or derived from dictionaries[33].

A. **Named Entity Recognition (NER)-** Recognising named entities, such as an individual's name, business organization, or location, is named entity recognition (NER).

B. **Analysis of Sentiment-** Analysing sentiment is the strategy that is utilised the most frequently in NLP. Analysis of sentiment is instrumental in contexts in which individuals communicate their thoughts and opinions, such as in the case of customer surveys, online reviews, and comments on social media platforms. The easiest way to interpret the findings of sentiment analysis is to categorise them as positive, negative, or neutral. In more complicated scenarios, the result may be a numerical score that can be segmented into required categories. With the assistance of sentiment ratings, we also find it helpful to determine which parts of the reviews are the most positive or the most negative.

C. **Text summarization-** Most of the time, this is used in news articles and academic papers. Extraction and abstraction are the two primary strategies that can be utilised when summarising a text. Extraction techniques result in summaries, which are created by removing specific sections of the text. Abstraction methods summarise by generating new text containing all the essential information when combined with the original text. Many algorithms, including LexRank, TextRank, and latent semantic analysis, are available for text summarization.

D. **Aspect mining-** The "aspect mining" process identifies many text facets. It extracts all of the data from a text when coupled with sentiment analysis. Using parts of speech tags is one of the simplest aspect analysis methods. The output accurately captures the complete intent of the text when aspect analysis combined with sentiment analysis is applied to sample text.

E. **Word Embeddings-** Word embeddings are a means to represent words in a document quantitatively. Similar words should have representations that are similar to that one. Words are depicted in contemporary theories as actual vectors. The length of each word vector is the same, but each vector's value varies. The difference between the two vectors shows how similar they are.

1.7 Word Embeddings Techniques

Word embedding is an algorithm which depicts words as vectors with precise values within a specific matrix field. This is accomplished by using a method known as word embedding. *Word embedding* is the method that is utilized to achieve this goal. The word "vector space representation,"[11] which is used to describe the procedure, can be considered an extension of the approach that was explained because it can be viewed as an extension of the phrase. Because every word is assigned to its unique vector, and the parameters of the vector are acquired in a way comparable to that of a neural network, this method is commonly used in deep learning.

I. **TFIDF-** Frequency of terms and frequency of inverse documents are two areas frequently using tf-idf weights. A statistical metric known as this weight is applied to a corpus, which can be considered a collection of documents to

evaluate the significance of individual words. Search engines frequently use different iterations of the tf-idf weighting system as a fundamental tool to score and rank the relevancy of documents based on a user's query.

II. **TF-** The Term Frequency (TF) algorithm determines how often a word is used in a text and counts those occurrences. Because the size of every record is unique, it is feasible that a particular term will appear much more frequently in a longer document than in a shorter one. This is because longer documents include more information than shorter ones.

TF(t) = (the total quantity of the word t in the corpus)/(total quantity of words in the corpus).

III. **IDF-** This is a metric used to gauge a term's significance. For determining TF, each term is given equal weight. It is acknowledged that some words, such as "is," "of," and "it," may appear numerous times without being given much weight. The formula below must be calculated to balance the frequent terms while exaggerating the rare terms:

IIDF(t) = log e (the total quantity of corpus divided by the number of corpus containing the word/term t).

IV. **Bag of Words(BOW)-** One widely utilised method is the bag-of-words method in information retrieval and natural language processing (IR). When training a classifier, methods for document classification frequently use this feature; the frequency with which each word appears is one of the features utilised in the training process. Pre-processing the data is the initial stage. The text should be changed to lowercase, and all non-word characters and punctuation should be eliminated. Finding the text's most frequently occurring words is the next stage. Each sentence must first be tokenised into words, followed by a vocabulary definition, before the word's frequency is counted. The model is then built after that. A vector is constructed to establish a word's frequency. If a word is frequently used, it is set to 1; otherwise, it is set to 0.

V. **Word2Vec-** The efficiency of Word2Vec can be attributed to its capacity to construct vectors consisting of related words. When provided with a sufficiently

extensive dataset, Word2Vec can make accurate predictions regarding the precise position of the word. Combining these estimations with other words discovered in the corpus results in the production of word associations. Some words, such as "King" and "Queen," for example, have a pronunciation that is very close to that of another word. By carrying out algebraic operations on word embeddings, you can obtain a reasonable estimate of the degree to which words are similar to one another. A significant number of the accomplishments that the word2vec programme has achieved can be attributed to the CBOW architecture and skip-gram designs. Their contributions were crucial to the program's success. The model was developed to estimate the probability that a specific term will occur when an input word does. When compared to the CBOW model, thinking of the skip-gram model as its antithesis makes perfect sense. The architecture in question allows for accurate predictions regarding the words before and after the current word by using the input word as a point of departure.

VI. **BERT-** BERT considers the text that immediately surrounds the text it is analysing to construct the context. This allows it to better assist computers in understanding the meaning of words that may be unclear. One deep learning model that employs transformers is BERT, an acronym for "Bidirectional Encoder Representations from Transformers." Within the Transformers framework, there is a connection between each output element and every input element. The weights assigned to these connections are determined dynamically based on the relationship between the two sets of characteristics. Additionally, there is a connection between every output element and every input element. In the past, language models analyze text input sequentially in only one direction. This was because the simultaneous interpretation was impossible. BERT is one of a kind due to its ability to read in both directions simultaneously. This capacity is known as bi-directionally, and it was made possible thanks to the development of Transformers.

VII. **RoBERTa-** Facebook came up with RoBERTa, a strategy that retrained BERT using improved training methods. RoBERTa is a robustly optimised version of the BERT strategy. Additionally, dynamic masking has been implemented to change the hidden token as the training epochs progress. Additionally, it was

found that having a more significant number of people in each training batch was more beneficial.

VIII. **DistilBERT-** DistillBERT uses a technique known as distillation, similar to Google's BERT, replacing the massive neural network with a more manageable one. Once trained, a smaller network can approximate the complete output distributions of a more extensive neural network. This resembles posterior approximation in specific ways. Kulback Leiber divergence, a crucial optimisation function for posterior approximation in Bayesian Statistics, has also been applied here.

IX. **XLNet-** With the help of XLNet, a generalised autoregressive language model, text sequence representations can be learned unsupervised. To overcome the drawbacks of AE, this model applies modelling strategies from BERT's Autoencoder (AE) models to AR models. XLnet has expanded the Transformer-XL model. Via the use of an autoregressive technique, it learns bidirectional contexts. By the transformer architecture with recurrence, the auto-regressive language model XLNet produces the joint probability of a series of tokens.XLNet implements permutation language modelling to enhance the training process. This model predicts each token but does it in an arbitrary sequence. The model can learn bidirectional relationships, and as a result, it can better manage dependencies and relations between words.

X. **GPT-2 /GPT-3-** Unsupervised deep learning transformers power the GPT-2 language model, which predicts a phrase's next word(s). GPT-2 and GPT-3 are enhanced versions of GPT. Its ability to handle more specialised issues makes GPT-3 more robust than GPT-2. It is well known that the GPT-2 performs poorly when given tasks in specialised domains such as music and narrative. GPT-3 can now perform additional tasks, such as question-and-answer sessions, essay writing, text summaries, language translation, and the creation of computer code. Even though it is still considered a language prediction model, a sequential text prediction model might be a better way to describe it. Because of the enormous amount of data used to pre-train GPT-3, its algorithmic framework has been recognised as among its category's most cutting-edge examples. This recognition comes from the fact that it is among the most advanced examples of its category.

After taking in information from the user, the GPT-3 system uses the field of study known as semantics to comprehend the meaning of the language being used and then produces sentences based on that information. The result should be a phrase that makes perfect sense to the person using it.

The ability to detect fake reviews increases consumers' confidence in online reviews. Consumers are empowered to make well-informed purchasing decisions grounded in authentic experiences through the filtration of deceptive content. Inauthentic customer feedback has the potential to damage the standing of an organisation and discourage legitimate patrons. Enhanced customer trust and loyalty result from the positive online reputation that detection assists companies in preserving. Consumers and enterprises can benefit from an environment that is trustworthy and conducive to good health online. By addressing the issue of fraudulent reviews, platforms can bolster user trust and guarantee the credibility of their reviews section.

Using user reviews to train and validate various fake review datasets depends on data sources, review volume, and filtering methods. Include consumer reviews from product review websites, app stores, social media platforms, and possibly internal user feedback systems in the collection. This function lets you capture many writing styles and formats for assessment. Use sentiment analysis to determine each review's favourable, harmful, or neutral attitude. This can help identify inflated reviews that may be fake—added sentiment to the false review detection model. Review sentiment scores that are exceptionally favourable or negative and significantly different from the projected distribution may be noted for additional review. Find commonly used keywords or phrases that indicate fake reviews, such as "amazing," questionable promises of perfection, or generic comments without product or service information. A comprehensive model for detecting false reviews can be created by integrating several data sources, rigorous validation methods, and sentiment analysis. This model can identify fake content and improve online safety for individuals and businesses.

Word Embedding Techniques for Text Classification			
	Types	**Pros**	**Cons**
1. Traditional	**Count Vector**	Real text data can be processed with count vectors, which will then yield precise word counts of the data.	Not provide any semantic information and could not compare the content of two
	TF-IDF	I. Insist on the significance of a specific term by pointing out how infrequently that term appears in all of the publications. II. Quickly determine the degree to which two corpus are comparable to	I. unable to capture the semantics or the co-occurrences in the various documents. II. unable to distinguish between singular and plural forms of
2. Static	**Co-occurrenceVector**	Able to preserve the semantic connection between the	Substantial quantity of Memory is required.
	Word2Vec	I. Map target word to context phrase to label unlabeled corpus. II. Non-linear connections are not implied.	I. Avoids global data. II. Sub-linear relationships aren't specified.
	Glove	I. Glow uses information on the local context of words in addition to data from all across the world (word co-occurrence). II. Ability to deduce semantic connections. III. Word vectors are capable of understanding sub-linear relationship	Requires more memory in order to store more information.
	Fast text	I. Word fragment-based and can handle unseen words. II. Ability to vectorize words not in OOV	Does not contribute any new contextual information.
3. Contextualized	**ELMo**	Several word embeddings to be used for a single word.	Since ELMo cannot benefit from both left and right contexts at once, it is shallow bidirectional.
	GPT2	GPT-2 predicts words by analysing sentences. The likelihood score can predict over ten words.	GPT-2 requires heavy computing and may generate erroneous results because it is trained using data from millions of websites.
	BERT	Easily detect the contextual relationships between words and subwords. The sentence's blank word can be assumed.	BERT only handles short sentences.

Table 1.4: Word Embedding Techniques

1.8 Deep Learning over Machine Learning

One of the complex tasks of opinion research and sentiment analysis in today's modern world is locating fake reviews on social networking platforms and fraudulent product reviews on online shopping sites.

Approaches based on machine learning can analyze and automate the process of finding fake reviews on social networking platforms and websites that sell products, and they can subsequently designate those reviews as spam. **Machine Learning** enables computers to learn independently without being explicitly taught. Machine learning is also abbreviated as ML, which stands for machine learning.

Deep learning algorithms spot such reviews by examining patterns and relationships in vast research databases when applied to fake reviews. Deep learning can help us understand patterns and relationships in enormous datasets, which is one of the most significant benefits of using this technique to detect bogus reviews. DL algorithms may learn to distinguish between genuine and false reviews; huge review collections should be used for training.

These methods may be classified as genuine or fraudulent and can be applied to future studies. Because of this, deep learning algorithms may be pretty efficient at identifying fake reviews, especially when compared with basic ML approaches.

The capability of deep learning to deal with high-dimensional information is another advantage it offers in identifying fake reviews. In the case of false reviews, DL techniques could be taught in vast review databases, which can comprise many aspects.

Deep learning models for detecting fraudulent reviews provide practical tools but can raise ethical concerns about user privacy. Large amounts of user data, including ratings, reviews, and maybe profiles, are usually required to train deep learning models. It is advisable to erase user data before feeding it into the model to prevent privacy threats. Deep learning algorithms that are taught on biased data have the potential to continue and reinforce those biases in their predictions. Identifying fake reviews might result in unjust or misleading categorizations, harming particular businesses or groups of users.

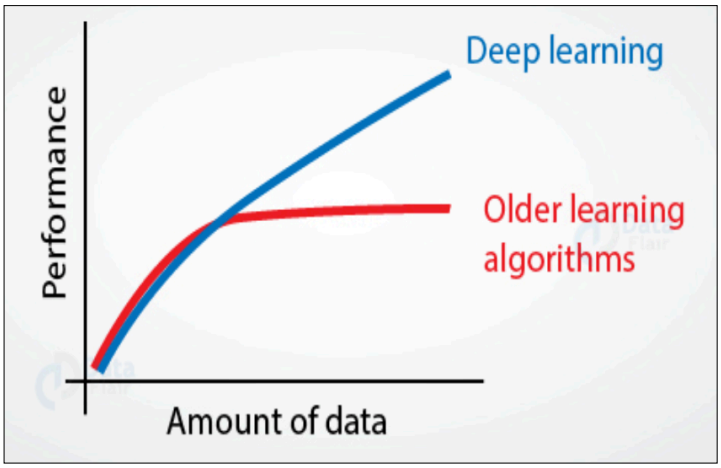


Figure 1.4: Deep Learning over Machine Learning

1.9 Problem Statement and Motivation

When making purchases online, customers have come to rely heavily on online reviews of goods and services to assist them in making selections about what to buy. Because of this, customers, producers, and merchants can all benefit from the knowledge provided by product reviews because it influences their purchase decisions. Customers discuss their experiences and provide word-of-mouth information regarding products, including their durability, functionality, and longevity.

The increase in internet-based platforms has augmented the resources available for gathering consumer feedback about the calibre of the products and services obtained. Many customers engage in online discourse to exchange opinions and experiences regarding various service-related matters.

As a result, businesses are spending more resources monitoring, analysing, correcting, and promoting their online reputation. The last few years have increased the prevalence of deceptive review spam, consisting of fabricated reviews intended to give the impression that actual customers wrote them. It is more probable that reviewers needing more familiarity with the goods or services under consideration will generate fabricated, excessively pessimistic, misleading, and fraudulent evaluations. Spammers disseminate fake reviews to either discredit or promote a

particular brand or product, with the end goal of convincing customers to make purchases from that brand.

This research is crucial because it helps identify spam user reviews in online service portals, a problem that exists. This research's findings will provide theoretical and practical contributions to the growing problem of spam reviews on various goods and services once they have been analysed. In addition, the findings would strengthen the current research on the methods used by multiple companies and organisations to identify spam reviews. This ensures that consumers are provided with reliable reviews of products sold through online marketplaces and that these consumers may be interested in purchasing.

1.10 Research Objectives

OBJECTIVE 1:	To study and evaluate the performance of existing models for Analytics on Online Social Network and E-Commerce sites.
OBJECTIVE 2:	To extract different patterns of unstructured data from Online Social Network pages and E-commerce sites through Analytics tools by examining customer reviews.
OBJECTIVE 3:	To identify the various parameters for recognition of fake reviews.
OBJECTIVE 4:	To propose a Fake Review Authentication model based on extracted patterns for customer reviews.
OBJECTIVE 5:	To evaluate and compare the performance of proposed model with existing models.

Table 1.5: Research Objectives

So, we conclude that electronic commerce platforms and Online Social Networks (OSNs) have extensively incorporated online reviews as an essential component. These entities substantially impact consumer behaviour, influencing purchasing choices, brand image, and user confidence. Nevertheless, many fake reviews gravely threaten these platforms' integrity, which are intentionally deceptive or misleading. Fraud reviews can manipulate consumers' opinions, damage brands' reputations, and impede equitable competition in the digital marketplace. To tackle this obstacle and build trust in the digital age, it is imperative to implement efficient detection techniques that guarantee a secure digital atmosphere where authentic user experiences can inform judgments.

Chapter 2

Review of Literature

2.1 Research Questions and Process of Searching

2.1.1 Questions To Consider For Research

To initiate the review planning process, we formulate four study inquiries that align with the intended objective. The format in which the queries have been presented is as follows:

• **Research Question No. 1-** What traditional and advanced strategies are used to identify fake reviews?

• **Research Question No. 2-** Which word embedding technique is suitable for detecting fake reviews?

2.1.2 The Search Procedure

Utilising a comprehensive assortment of databases is critical for ensuring that the study's scope aligns with the objectives and optimising the likelihood of locating pertinent articles. As a consequence of this, searches for primary studies were conducted using the following three bibliographic databases:

• ScienceDirect,

• IEEE

• Acm

The search criteria employed for this research study comprised the following terms:

(detect," "opinion spam" or "fake reviews"); (machine learning," "supervised," or "unsupervised," or "deep learning"); and (detect," "opinion spam," or "fake reviews").

2.2 Criteria for Inclusion vs Exclusion

The criteria used in a review need to be precisely outlined to reduce any potential for bias. Sometimes, the chosen studies may be outside the research question or the project's goals. Having selection criteria eliminates the need for such concerns.

2.2.1 Inclusion Criteria

• Include research mainly from the period beginning in January 2018 and ending in 2022

• Only studies written in the English language are considered valid.

• Research Articles and Studies that Have the Potential to Answer Either of the Research Questions Presented in Section 2.1.

• Only papers and articles previously presented at conferences or published in journals.

2.2.2 Exclusion Criteria

• The researchers didn't consider any of the papers that were published more than once.

• Research that isn't relevant to the question being asked in the research.

• The full texts of the studies cannot be accessed.

• Documentation as a conclusion to the study or a memoir.

2.3 Results and Answers

Machine learning techniques have been increasingly employed in various domains, including the diagnosis of diseases, face identification, recognition of voices, font recognition, and fraud detection, among others.

Studies have also been undertaken in recent years on using machine learning to combat spam, an issue spreading across various internet applications such as SMS, email, and blogs.

2.4 Research Question No. 1: What methods and strategies are used to identify fake reviews?

Jindal and Liu [33] created an algorithm for supervised learning that may detect false reviews by examining duplicate reviews. This allowed them to identify bogus reviews. The proposed framework has two distinct stages. At the initial research

stage, unigrams and bigrams served as the project's features, and the Naive Bayes, Random Forest, and Support Vector Machine classification techniques were used. The classification methods' performance was enhanced using two ensemble approaches in the project's second phase. Stacking and voting were the two strategies in question. The results on the AMT dataset indicated that the ensemble approaches produced better results than the Naive Bayes random forest and SVM classification algorithms. This was proved because the ensemble techniques delivered superior classification accuracy. These findings were demonstrated by analyzing the degree of accuracy displayed by each algorithm's predictions. Using straightforward feature algorithms in conjunction with ensemble methods can result in a considerable improvement in the accuracy of the detection of fraudulent reviews. On the other hand, if multiple checks are dishonest, the information may not be reliable.

Lin et al. [34] established a classification model to detect fake reviews in a cross-domain context. This model was based on a Sparse Additive Generative Model (SAGE) developed based on the Bayesian generative model. The model amalgamates a generalised additive model and topic modelling. They employed linguistic question and word account (LIWC), POS, and unigram approaches as features to identify fake reviews in cross-domains. The suggested model might consider various facets, such as fake versus accurate and positive versus negative. To evaluate the proposed model, they used the AMT dataset, which included reviews for three types of businesses: restaurants, doctors, and hotels. According to the experiment's findings, the unigram categorization accuracy was 65%. Using unigram, we determined with 76.1% accuracy two different class classifications (Turker reviews and Employee reviews). The accuracy using unigram, POS, and LIWC separately for the restaurant domain was 77%, 74.6%, and 74.2%, respectively. The accuracy on cross-domain while employing unigram, POS, and LIWC independently with the Doctor domain was: 52%, 63.4%, and 64.7%, respectively. Nevertheless, the developed model could have more successfully captured the semantic information in the sentence.

Hernández-Castaeda et al. [35] studied the effectiveness of employing SVN (Support Vector Network) in classification tasks to detect false reviews in one, mixed, and cross-domains. They utilized the LIWC, Word space model (WSM), and latent Dirichlet Allocation (LDA) methodologies as feature extraction strategies.

They used the DeRev, OpSpam, and Opinions datasets to test the suggested approach and determine its efficacy. When the results were compared to those of the previous works, it was found that a combination of WSM and LDA achieved the best results in one domain, with an accuracy of 90.9% on the OpSpam dataset, 94.9% on the DeRev dataset, 87.5% on the Abortion dataset, 87% on the Best Friend dataset, and 80% on the Death Penalty dataset. There was also an improvement compared to the accuracy achieved by the Naive Bayes classifier, which was 76.3% in a mixed domain. They used the dataset for testing and mixed the remaining datasets for training, so the performance was good in one domain and the mixed domain, but it was terrible in the cross-domain test. A deep neural network is probably more appropriate for enhancing fake review identification in a cross-domain by increasing the learning presentation.

Sedighi et al. [36] recommended using decision trees to detect fake reviews. They chose and analyzed features using traditional approaches. Considering data correlation when choosing crucial components could improve the model.

Khurshid et al. [37] proposed a content- and first-feature-based supervised machine-learning technique to detect fake reviews. Five classifiers—Naive Bayes, Random forest, JRip, AdaBoost, and J48—classified reviews in the proposed model. Using real-world data [8], the AdaBoost method with combined features outperformed other classifiers with 73.4% accuracy. Primal factors increase performance significantly. However, an imbalanced dataset may have improved the proposed model.

Khurshid et al. [38] proposed an ensemble learning approach to identify bogus reviews based on their attributes. Earlier in his study, the author discussed that Tier 1 used Discriminative Multi-nominal Naive Bayes, a Support Vector Machine package, and J48, whereas Tier 2 used Logistic Regression for accuracy. The two-tiered model is suggested. The following feature selections extracted structural and language features: Vector space greedy stepwise. Chi-squared was used to calculate an attribute's worth. Particle swarm optimization explored feature space; Cuckoo Search analyzed attribute space and Greedy stepwise analyzed vector space. They tested the proposed model using real-world and semi-real-world data. The Ensemble and Hybrid Modelling techniques are two prevalent methods used to construct more

accurate models for identifying fake reviews. When multiple models' predictions are merged in ensemble modelling, a more dependable and precise model is produced due to the process.

Singhal et al. [39] suggested a system that utilises decision trees to identify fake reviews. They selected appropriate features by employing standard methods of feature selection, and then they evaluated those features. One way using the presented model might be improved by considering ensemble parts of the model(WSEM-S) that correlation between the data when considering which components are essential.

Deshai et al. [40] suggested a supervised machine-learning strategy to detect fake reviews based on content features and first features. The proposed model utilized the following six classifiers to categorise the reviews: Naive Bayes, Random forest, JRip, AdaBoost, and J48. The findings obtained from a dataset derived from real-world data [8] demonstrated that the AdaBoost algorithm with merged features performed superiorly to other classifiers, achieving far better accuracy. In addition, utilising Primal elements considerably impacts the performance improvement achieved. The proposed model, on the other hand, could have performed better when applied to an imbalanced dataset.

Khurshid et al. [41] built upon their earlier work and suggested an ensemble learning model to identify fake reviews based on a selection of their characteristics. Tier 1 utilized three classifiers (Discriminative Multi-nominal Naive Bayes, a library for Support Vector Machine, and J48), while Tier 2 utilized a Logistic Regression classifier to introduce an accurate result. The suggested model is comprised of two tiers. In addition, they extracted structural and linguistic features using the feature selections listed below: Greedy stepwise was carried out in vector space. Chi-squared was used to evaluate an attribute's worth by calculating the Chi-Squared statistic value. Particle swarm optimisation was used to explore the feature space, Cuckoo Search was used to analyze the attribute space, and Greedy stepwise was carried out in vector space. They tested the suggested model using data from real-world situations and semi-real-world situations. The results of the experiments demonstrated that the chi-squared feature has a significant role in boosting the performance of the proposed model, which achieved an accuracy of 84.1% on the dataset consisting of Yelp restaurants and 81.7% on the dataset consisting of semi-

real data. However, if the chi-squared feature were incorporated into the deep learning model, the performance of the suggested model may be significantly enhanced.

Cardoso et al. [42] examined the differences between content-based classification models to determine whether or not specific data characteristics shift over time. Yelp datasets revealed that the performance of the models deteriorated over time [8]. Spammers attempted over and over again to get through the spam filter. In applications based on the real world, the most recent reviews contain features that need to be confirmed by a model trained with reviews from earlier in the process. In addition, the performance of the models worsened over time.

Additionally, review direction affected approach effectiveness. They advised using a method adapted to each polarity. They also found that different products and services affected technique performance. They offered a different model for each product or service.

Sánchez-Junquera et al. [43] suggested a character n-gram model for detecting fake reviews. They used Naive Bayes and a support vector machine. "Death penalty," "Abortion," and "Best Friend" domains were used to test the model. The recommended model outperformed SVM with LIWC, LDA& words, and Deep syntax & expressions in detecting fake reviews. Other methods yielded better results. Combination features may increase classification model performance.

Using basic ML models, **Mani et al. [44]** developed a two-phase bigram features model for unigram-based fake review detection. The stacking and voting ensemble improved the classification model's performance in the second development phase. The gold standard dataset showed that the Naive Bayes method has the highest starting accuracy (87.21%). The stacking ensemble outperformed voting with 87.68% accuracy. Even if deep learning outperforms the suggested model, the ensemble technique showed the importance of searching for fraudulent testimonials.

OneReview was introduced by **Nilizadeh et al. [45].** OneReview isolates odd business profile updates on many review platforms to discover detrimental actions. OneReview used change point analysis on each website review. They then checked the change point for discrepancies. The Random Forest classifier found them after the shift point analyzer indicated fake reviews. The approach was tested on the TripAdvisor and Yelp Data Challenge datasets. The suggested method recognised

fake reviews with 97% accuracy for full features and 86% for textual features. Unfortunately, the change point analyzer analyzed time series data in one month, which may delay OneReview categorization and review publication.

Li et al. [46] presented hybrid supervised machine learning spammer detection. Since reviewers' behaviour is temporal, they used labelled hidden Markov to identify spamming by single reviewer posting time. Multi-hidden Markov was applied to Co-bursting posting signals and behaviour. Co-bursting detects spammers. These methods used real datasets without metrics to test the model.

Sánchez-Junquera et al. [47] provided a cross-domain fake review detection adaption method. Co-occurring Entropy and a mismatch technique were used to discover and hide domain features. Naïve Bayes classifiers missed cross-domain fake reviews in the gold standard dataset.

The authors [48] acknowledged a concept drift problem in false reviews, where reviews change over time, but did not address it. The authors used statistical machine learning and benchmark concept drift detection to explore and prove their assertion. The authors employed statistical machine learning and benchmarking concept drift detection to demonstrate their conclusion. They tested the classifier on four Yelp datasets. The classifier's performance dropped due to fake reviews' changing properties. They also found that concept drift negatively affects classification performance and prediction algorithm performance. This study emphasises the need for counterfeit review detecting systems.

The authors [49] suggested using content, language, and rating to detect fake reviews. The gathered attributes are entered into SVM, NB, RF, and MLP machine learning classifiers to detect if the review is fraudulent. They used Yelp.com data to test the model. Review inconsistency features may improve fraudulent review identification, according to experiments. However, the model can work with little data and improve performance by combining word embedding representation and deep learning.

Yao et al. [50] proposed a review content and reviewer feature-based ensemble fake review detection algorithm. Grid search and resampling were used to tackle unbalanced data. Each classifier receives the retrieved features separately. Finally, majority voting and stacking improve model performance. The suggested model did

not outperform state-of-the-art approaches in Yelp dataset experiments. The model requires higher time complexity. Data classification projects, especially natural language processing ones, benefit from neural network techniques. Deep learning's most representative neural networks can quickly extract beneficial data properties. Word embedding and deep learning can extract text semantics. The Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) have been extensively studied to detect fake reviews.

Rout et al. [51] further emphasise the need to use feature engineering methods to detect fake reviews considering all three types. Their research goal was to take all extracted data, apply supervised, semi-supervised, and unsupervised learning methods, evaluate their outcomes, and choose the process with the highest prediction accuracy.

Siagian et al. [52] used word and character n-grams to detect fake reviews.

Martinez et al. [53] took a set of unique attributes based on the reviews' polarity and concentrated on the text's content. This combination's various properties make machine learning methods challenging to use. PCA segregated dominant feature attributes from non-essential feature attributes to reduce feature attribute size. Applying Machine Learning classifiers to label testing data provides the final data to be learned.

Taneja et al. [54] suggested an Ensemble model study supervised data categorization. Labelling the Cloud Armour dataset was the strategy. Restricting review counts, service counts, and probability of collusion feedback factors added these labels. These limits allowed classification model-based supervised machine learning on the labelled dataset.

Wang et al. [55] propose multi-feature fusion, which combines sentiment analysis, review text features, and reviewer behaviour features collected with a related approach (Doc2vec for text representation as a pre-processing step). Next, they choose the most accurate classifier from seven in the sample labelled dataset to classify new reviews. Finally, the result of this step is added to the initial models.

Aiyar et al. [56] use techniques such as Random Forest, SVM, and NB, in addition to proprietary heuristics like character n-grams, to identify spam reviews and devise strategies to combat them.

Jamshidi Nejad et al. [57] showed that Decision Tree and AdaBoost could detect false reviews using a new set of data features from text normalisation and part-of-speech tagging.

Wang et al. [58] propose two characteristics and categorise data using supervised machine learning. Two additional semantic features—readability and theme—were investigated. These features include readability. They indicate that fake reviewers and reviewers detect bogus reviewers and sites. The Automated Readability Index (ARI) and Coleman-Liau Index (CLI) were introduced to assess review readability. Reviews guided this. They gave behavioural factors such as restaurant number (RN) and date interval (DI), one of the reviewers' alternative perspectives. N-gram characteristics from natural language processing technologies, such as unigrams and bigrams, are also used to identify bogus reviews.

On the other hand, **Noekhah et al. [59]** illustrate the intra and inter-relationships among entities under the model, analyze the importance of features by calculating weights based on feature fusion techniques, and, as a result, determine the most practical combination of weighted features. This method can be used to illustrate the intra and inter-relationships among entities. Following that, the authors built a multi-iterative algorithm to update the spasticity. During the preprocessing stage, the noise was removed, and the text was standardised to be consistent with the structure and the data content. Afterwards, the information Gain (IG) and the TF-IDF are utilized for the feature selection. We select the highest potential value characteristics using well-known classifiers like SVM, NB, and DT. After that, they use "feature fusion" to determine which combination of features is the most useful and practical. Ultimately, they use a multi-iterative method to run a certain number of iterations and then determine spasticity based on the results.

Mewada et al. [60] discuss unsupervised fake detection as a supervised alternative to the traditional approach. In unsupervised machine learning, the strategy involves clustering the input data. Therefore, this approach does not require labeled data to detect bots and cluster accounts based on similarity rather than relying on feature values to categorise each account reported in two experiments illustrating this strategy. These studies show that this approach was successful in identifying bots. After the data has been preprocessed, k-means clustering is performed based on the semantic similarities shared by review texts to pick a diverse group of reviews from

which the effect of textual attributes on trustworthiness is evaluated. In the second study, multiple regression models were used to investigate the influence that review valence, attribute salience, and content concreteness have on the reliability of reviews.

Gao et al. [61] focus on a neglected developing topic termed "movie review" and propose a novel unsupervised spam detection model using an attention mechanism. It has been found that customers will transmit their thoughts towards various aspects of films in the reviews they write and that these sentiments can be gleaned from the statistical elements extracted from reviews. An attention mechanism is added to the review embedding, and a conditional generative adversarial network is utilized to learn users' review styles for various film genres.

Wang et al. [62] proposed an unsupervised network embedding-based strategy to simultaneously integrate direct and indirect neighbourhood exploration to learn user embeddings for more accurate spam reviewer identification. The authors of the study took this approach. After measuring the level of spammer agreement based on their direct co-rating associations, direct relevance is used to build a user-based signed network based on the immediate relevance of the user for embedding. Direct relevance is also used to determine the degree to which spammers agree with each other. At the same time, indirect bearing uses a truncated random walk to quantify the relevance of positive users to indirect information. The writers consider the Product Rating Proximity, the Product Time Proximity, the Category Rating Proximity, and the Category Time Proximity.

Ennaouri et al. [63] compared six of the most prominent semi-supervised learning algorithms with three supervised classification methods (SVM, NB, and RF). Semi-supervised learning algorithms include graph-based learning, self-training, co-training, multi-view learning, and low-diversity separation (LDS) generative approaches. [Graph-based learning] refers to learning in which a data set is represented as a graph. First, once the standard preprocessing task has been finished, they select 1000 top features that they believe to be the most accurate predictors by performing the chi-square test (with unigram and bi-grams) and considering the high F-value. After that, the classification is used, and its accuracy, precision, recall, and f1-score are analyzed and compared. In addition, to locate the hyperparameters that offered the best gain in performance, they utilized a Grid Search for each base

classifier using a semi-supervised technique. This allowed them to find the optimal hyperparameters.

Hassan et al. [64] evaluated semi-supervised (Expectation Maximisation) and supervised (NB and SVM algorithms) text mining models to detect false online reviews. Expectation Maximisation: Using the labelled dataset, a classifier is developed. The classifier is extracted from the unlabelled and labelled datasets and used to classify the unlabelled dataset again. The classification algorithm predicts the test dataset after training with the combined training set.

Recent years have seen uncontrolled spammer group detection of fake reviews. **Zhang et al. [65]** suggested context-dependent spammer group detection. The researchers collected labelled and unlabelled data by following the methods above. Frequent item mining (FIM) was used to identify spammer group prospects. Manual marking of positive spammers followed. After that, an expectation maximisation (EM) technique is used to incorporate an unlabelled dataset into a naïve Bayes classifier trained on a labelled dataset.

Neural network methods function well for data categorization, especially natural language processing. Deep learning's most representative neural networks can quickly extract beneficial data properties. Word embedding and deep learning can extract text semantics.

Li et al. [66] developed a convolutional neural network model for learning document representation to recognise deceptive spam views. The model was proposed using a word vector. A sentence-weighted neural network model represents each phrase and document in the review. The suggested model has two convolutional layers: the sentence layer, which composes the sentence, and the document layer, which converts the sentence vector into a document vector. CNN was influential across domains. CNN outperformed LSTM in mixed-domain tests.

Zhao et al. [67] identified fake reviews using a CNN approach that retains word order. Word vectors were created using word2vec using word order reserving instead of max pooling. Accessible output layers are created by concatenating pooling layer features. The AMT dataset and 10,000 data-annotated reviews were used to test the model. Experiments showed that the suggested model outperformed state-of-the-art approaches with 70.02% accuracy. CNN categorises brief text reviews better. CNN

trained faster, while RNN fared better on long-text tasks. Many people find the hand-annotated method laborious.

Wang et al. [68] used an attention neural network to improve the classification model by identifying false reviews based on language, behaviour, or both. The model measured emotional weight using language and behavioural trends throughout training. A multi-layer perceptron extracted behavioural information, whereas a CNN removed linguistic aspects. The emotional weight of linguistic and behavioural traits was then learned through attention. The Yelp dataset showed that the suggested model surpassed the state-of-the-art techniques with an accuracy of 88.8% in hotels and 91% in restaurants. The attention mechanism also boosts classification model performance. However, the model prioritised language features above behaviour features, which must be changed to detect fake reviews.

Wang et al. [69] constructed an unsupervised neural network model to detect fraudulent reviews using behaviour and content. This concept solves the cold-start problem when a new reviewer writes a review. CNN was used to model the review text's semantic content, which unigram and LIWC cannot convey. The model used behavioural and textual data to distinguish reviews. TransE can also represent network nodes and edges used to encode behavioural information. The recommended model outperformed SVM on the Yelp dataset with 65.4% hotel domain accuracy and 62% restaurant domain accuracy. Learning review embedding with encoded language and behaviour information works better. The model was not compared to other embedding or dimension reduction methods.

Zhang et al. [70] introduced the DRI-RCNN false review identification model utilising a word context-based recurrent convolutional neural network. The proposed model has four layers: a convolutional layer to train the vector to represent a word and a recurrent neural layer to learn right and left for a false and real context vector of a comment. The model was tested using AMT and Deception datasets. The proposed model outperformed state-of-the-art methods like LIWC and unigram with SVM, LIWC feature, and four n-grams with SVM, recurrent convolutional neural network, profile alignment compatibility method, sparse additive generative model, lexical production rules with SVM, convolutional neural network, and gated recurrent neural network with 82.9% accuracy on AMT datasets. The suggested

model also performed well on the misleading dataset, with 80.8% accuracy. However, the paradigm requires significant time.

Li et al. [71] proposed embedding to tackle the "cold start" problem in detecting bogus reviews. This strategy would affect user reviewers' social interactions. The model incorporated a user's inferred review rating. This depiction evaluated user behaviours and item-user relationships. Item, rating, review, and user embedding layers comprise the model. After adjusting its success rate for a specific parameter, they embedded a co-occurrence-based user behaviour. The user-item network was built by evaluating activities that included user/item social relations based on random walk context information. CNN embedded text using CBOW.

Li et al. [72] developed an unsupervised model for fake review identification's cold start problem. This model is built on their previous work. They used emotional connections re-weighting instead of content and social ties to evaluate behavioural representation. Yelp NYC and Yelp Zip datasets tested the proposed model. The model scored 60% in the hotel domain and 70% in the restaurant sector. The proposed model did not outperform the current best technique and ignored review text features that could increase classification model performance.

Ren and Zhang [73] used gated recurrent neural networks to learn document representation for fraudulent review detection. CNN's word-to-sentence sentiment analysis was best. Gated RNN and attention-created document representation for fraudulent review detection. They used data from doctors, hotels, and restaurants. Hotel and restaurant domains outscored doctors due to unclear terminology. They solved this problem with neural feature discrete logistic regression. They concatenated neural components into discrete elements—81.3% for hotels, 87% for restaurants, and 76.3% for doctors—before the SoftMax layer. Hotel reviews trained the classifier to 83.7% accuracy in restaurants and 57.3% in doctors. The recommended model outperforms RNN, CNN, GRNN, and Bi-directional GRNN in one and cross-domain. Features may improve classification model performance. Unfortunately, the proposed model needs time complexity.

RNN can process sequential data utilising internal memory for the input sequence. RNN can store long lines. However, RNN can only run a few steps before encountering expanding or vanishing gradients. Due to RNN restrictions, researchers

have developed additional models. These models are LSTM, Gated Recurrent Unit (GRU), Bidirectional, Stacked, and Attentive LSTM.

Wang et al. [74] used a dictionary-based long short-term memory recurrent neural network to detect spammers. A multilayer perceptron was created with an input layer, an LSTM layer, a hidden layer for dimension reduction, and an output layer for one neuron. The neuron's value determines the reviewer's spamminess. The dataset came from Taiwanese moblil01.com and product reviews. Data were annotated using internal secret materials. The suggested model found that long short-term memory detected false reviews better than SVM at 89.4%. Long-term memory makes LSTM better than RNN. The model did not compare to other neural network approaches. The proposed model only considered text and neglected metadata and behavioural features that could increase performance.

Liu et al. [75] built the bidirectional LSTM model to learn the document-level representation of reviews to recognise fake reviews based on their features to overcome the RNN's shortcomings. The proposed approach combines feature representation (POS), first-person pronoun characteristics, and document representation (word embedding (Glove)). Performance on the three-domain AMT dataset determined the model's validity. The recommended model outperformed state-of-the-art methods, including paragraph average, SWNN, SWNN+POS+I, BiLSTM, and basic CNN+POS+I. The suggested model outperformed SWNN, Deep CNN, CNN-LSTM, and CLSTM with 83.9% accuracy. The hotel domain had an accuracy of 83.9%, the restaurant domain 85.8%, and the medical sector 83.8%, beating the basic techniques. The model accuracy proves that checking for first-person pronouns is crucial to detecting dishonest reviews. Unfortunately, the proposed solution requires extensive computer resources.

Jain et al. [76] recently established hierarchical CNN-GRN deep learning algorithms, and Multi instant learning (MIL) methods were proposed to manage review lengths in false review detection. Three-layer CNNs extracted localized n-gram features. Instead of CNN, GRN learned semantic connections between extracted characteristics. The input text is often split into multiple cases, and the last sample is deleted if the word length is less than fifteen. The model was tested on the four-city dataset, Yelp Zip dataset, Deceptive Spam Corpus, Drug Review dataset,

and Large Movie Review dataset. MIL and CNN-GRN outperformed CNN and RNN on all datasets. Unfortunately, the paradigm only works for short material.

Zeng et al. [77] recently developed an ensemble technique to detect fake reviews based on review structure. They found that false reviews were more emotional than real ones. Many bogus reviews begin or end with similar sentences, and the paragraph's opening and closing statements portray a better variety of emotions than its content. The model encodes the review's beginning, middle, and end with three bidirectional LSTMs. These four representations are combined to identify bogus reviews. Self-attention united the three local representations. The attention strategy combined the two representations. The AMT dataset demonstrated that the proposed model outperformed SWNN and SAGA in one area (Hotel, Doctor, and Restaurant) with 85.7%, 84.7%, and 85.5% accuracy. The model's mixed domain accuracy was 85.7%. 83.4% correct. However, in cross-domain, the proposed model scored 71.6% in restaurants and 60.5% in medicine.

Dhamani et al. [78] suggested neural networks and transferred learning for social media misinformation. They proposed a character-level convolutional neural network ensemble with long-term and short-term memory. The approach applies labelled data from one domain to another. Transfer learning is also promising for distinguishing by-product behaviours. Using n-grams, the proposed model was simpler.

Aghakhani et al. [79] developed a semi-supervised false Generative Adversarial Network (FakeGAN) model to detect false reviews without datasets. One generative model and one discrimination model generate samples with the same distribution. Two differentiators solved the generator's convergence problem, making it far more powerful. The first can spot fake reviews. The second one distinguishes LSTM generative model reviews from fake reviews. Maximum likelihood estimation trains the generator to recognise fake reviews. The AMT dataset showed that the suggested model was 89.2% accurate. The proposed model did not outperform the state-of-the-art technique. Due to GAN's stability, the proposed model's hyper-tuning is difficult.

You et al. [80] used deep learning to solve the cold start problem in fake review detection using domain-specific features. They proposed encoding objects, reviewers, reviews, and their properties, such as date, price range, and location. They also advised utilising a domain classifier to transfer skills between fields. The proposed technique had three layers: the first layer integrated a variety of attributes into the

model; the second layer captured the three relations (entity-entity), (entity-attribute), and (attribute-attribute); and the third layer implemented a domain classifier to capture domain correlation. The suggested model outperformed SVM in the hotel and restaurant domains, with 80% and 75.6% accuracy, respectively. Generative adversarial networks help solve cold-start concerns. The model was not compared to other embedding methods.

Tang et al. [81] developed a generative adversarial network model to handle the fake review cold start problem. New users without features get synthetic behaviour features automatically. First, three easily accessible functionalities for new and frequent users were extracted. Six regular user features were then extracted. A GAN generator generated synthetic behaviour traits using easily accessible features. GAN has six levels. Normalisation and easy-to-access features were obtained from the first three levels. The other three tiers turned readily available features into synthetic behaviour traits. After training with GAN's discriminator, the generator is applied to the new user to acquire synthetic behaviour features. Yelp Chi [8], which has two domains (Hotel and Restaurant), was used to test the approach. The suggested model outperformed the state-of-the-art approaches with 83% hotel accuracy and 75.7% restaurant accuracy. The combined features improved classification model performance. The approach may have better detected false reviews in cross-domain applications.

Wang et al. [82] introduced a reviewer-product spam detection methodology. This approach was created to identify fake reviews without relying on reviewers' expertise. They made a 3-mode tensor using two things' relations. RESCAL tensor factorisation algorithms automatically learned product and reviewer vector representations. Finally, a support vector machine classifier is applied to the concatenated and final review expression. The suggested model outperformed the state-of-the-art approach with 85.9% hotel domain accuracy and 87.8% restaurant domain accuracy. The proposed model found that reviewer-product relationships improve classification model performance.

Wang et al. [83] created a multi-dimensional time series model to detect one bogus review. Combining trustworthiness and expertise is a unique measure that assesses reviewers' credibility. The ranking method grouped spammers into dimensions to discover aberrant time series features. Any bogus review in the time series reduces

the window size. The 408,469 evaluations from multiple sources showed that the proposed model was adequate for human assessment compared to the average RHR. They found various studies posted on different days between 2009 and 2010. The proposed model evaluated the procedure without recall, precision, the F1 measure, or accuracy.

Heydari et al. [84] developed a pattern recognition algorithm using metadata and rating deviation features to identify potentially false reviews made during reasonable periods. A time series analyses product review oscillations. Sliding windows record questionable periods and patterns. The presented method detected 86% of false reviews on Amazon.com datasets. Despite its benefits, the proposed technique focused on questionable periods rather than lowering costly computations during scoring. Hand annotation requires a lot of human resources, and adding metadata like an IP address can improve the suggested approach.

Li et al. [85] created a sentence-weighted neural network model (SWNN) to detect fake reviews. The suggested approach transformed each sentence into a document vector and related it to the weight. One reviewer's sentence stood alone. POS and First-Person Pronoun analysis was applied to verify the review. The AMT dataset, with hotel, restaurant, and doctor domains, was used to test the model. The unigram feature performed best in restaurants, with 78.5% accuracy. In the doctor domain, combining characteristics performed best with 61.5% accuracy. SWNN outperformed state-of-the-art algorithms on the mixed field with 80.1% accuracy [119], [169]. The hotel domain had 83.7%, the restaurant domain 87.6%, and the doctor domain 82.9% using the F1 score. It failed to forecast mix and cross-domain test results.

Noekhah et al. [86] created an unsupervised Multi-Iteration Network Structure to detect a single false review, a group of reviewers, and a reviewer using the reviews' behavioural and structural properties. The suggested model used inter- and intra-relationships (product, reviewer, and review) to extract features. The proposed model performed 98% using combination features, 74% using behavioural characteristics, and 69% using structural features on the Amazon.com dataset. They did not compare the proposed model to other methods to prove its utility. They did not employ all metadata features, which could increase classification model performance.

Yuan et al. [87] presented a hierarchical fusion attention network that learned product and user representations to detect fake reviews. A user-product multi-attention unit is presented to extract a sentence representing user-product features. Fusion attention units and orthogonal decomposition taught the user-product model. In conclusion, reviews are user-product relationships. TransH models, which embed knowledge graphs in vectors, encoded the product-review-user relationship. Mobile01 Review and Yelpdata assessed the model. The proposed model outperformed SVMs with content and behavioural features, Graph-based model (RSD), SpEagle, Tensor decomposition model (TDSD), Couple Hidden Markov model (CHMM), Spam2Vec, CNN-GRNN, SWNN, ABNN, and AEDA. The Mobil01 initial post dataset gave it 86.96% F1 and the Reply dataset 48.37% F1. On the Yelp Chi dataset, it scored 83.24% AUC, 84.78% on NYC, and 87.28% on Zip. The model showed that bogus review detection requires product and user levels.

Cao et al. [88] suggested the detection of deceptive reviews. The framework implicitly extracts the review semantics using fine-grained and coarse features. The retrieved features were learned via a coarse-grained concatenation that was comprised of two neural network layers and LDA. The use of deep learning allowed for parallel knowledge of fine-grained features. An SVM algorithm is trained to determine whether the review is genuine. Using a "gold standard" and a real-world dataset, the performance of the suggested model was shown to be enhanced in either one or both areas. LDA combined with Text CNN fared the best on one-domain and mixed-domain datasets. In addition, rougher features fared significantly better than their more refined counterparts.

To identify fraudulent reviews, a hybrid deep-learning strategy has been described [89] as a means of extracting review semantics. There was a discussion of three stages: During the first stage of the process, the review embedding was removed with the help of the Denoising Autoencoder and the Sentence Vectors Distribution Bags of Word. After that, to ensure the accuracy of the review, the representation of the features layering from the two models is concatenated and placed into a fully linked layer. On a dataset considered the gold standard, the recommended model's accuracy was 92.5%, better than contemporary methods' performance. Adding emotional factors will boost performance.

Guo et al. [90] advocated embedding occasional and stable links in graph neural networks to identify spammers. The parametric random walk technique extracted rare relations, while direct vectorised encoding modelled the stable connection—deep graph learning, introduced recently, simulated interaction properties. The model beat CNN, MLP, SVM, and LSTM on two real-world datasets.

Using pre-trained features from Global Vectors for Word Representation, **Archchitha et al. [91]** created a CNN model that can identify spam comments containing opinions. CNN was developed using this paradigm. To combine text and behavioural data, it was necessary to use three parallel convolution layers with varying filter widths. Word- and character-level features derived from previous research are taken from the written content and mixed with a feature set discarded by the convolutional layers to improve the model's performance.

Shahariar et al. [92] propose deep-learning spam-review detection. These include MLP, CNN, and LSTM. NB, KNN, and SVM were used to detect fake reviews. Finally, they compared deep learning and machine learning classifiers. Finally, 52% of the research employed supervised learning.

Budhi et al.[93] suggest 133 features: 80 content, 29 behaviour, and 24 product features. To correct minority class statistics, they over or under-sampled. 10-fold cross-validation tested MLP, LR, DT, CNN, and SVM classifiers. Parallel processing increased speed.

2.5 Research Question No. 2: Which word embedding technique is suitable for detecting fake reviews?

2.5.1 Traditional Word Embedding

The frequency method is the standard method for conventional word embedding. This method considers the entire document, determines the significance of unusual terms contained within it, counts the number of times each word appears, and also considers the frequency with which other words appear.

El-Din [94] devised an innovative technique for improving by using bags of words to address the problem brought on by the manual evaluation of words. This embedding examined the polarity of the sentiment and assigned a score to it based on the significance of the phrases in the sentence. The score was determined based on how

strongly people responded one way or the other. When selecting a person's disposition, the recommended method was accurate 83.5 per cent of the time.

A novel strategy for text mining was introduced by **Dadgar et al. [95]**, and it is based on embedding the TF and IDF to categorise news. The placement of the word inside the text served as the basis for determining how significant it was. Using the BBC and 20 newsgroup datasets, TF-IDF with the SVM classifier improved performance, obtaining a precision of roughly 97% and 94%, respectively.

Qu et al. [96] provided a novel TF-IDF strategy with an improved weight formula as an alternative to the TF-IDF embedding method, which only addressed document linkages and not characters' relationships. The novel TF-IDF and distance vector classifiers could generate a classification effect with an F1 score ranging from 80% to 90%; nevertheless, they might be improved to reach greater classification accuracy.

Tripathy et al. [97] conducted a study that compared and contrasted the findings of the SVM classifier with those of the Naive Bayes classifier. The authors utilized TF-IDF embeddings so that information provided in textual form could be translated into a format that could be represented numerically. When paired with proposed classifiers, it achieved an accuracy of 94%.

Enrquez et al. [98] developed a vector-based word representation using Word2Vec and a bag of words. The BoW method relies on using vectors to describe each document accurately. These vectors have a high degree of dimension as a result of the fact that each dimension corresponds to a word or group of words. BoW is a lexical representation based on frequency-based metrics that determine whether or not a word is included in a phrase. This representation generates massive vectors with very little information. Word embedding is a different strategy for turning text into numbers. This approach uses a vector space whose dimensions are lower than the vocabulary size to accurately capture the vocabulary terms. When both BoW and Word2Vec are used together, the accuracy of the representation is increased to approximately 75%.

Soumya et al. [99] proposed a BOW strategy using an F1 score that increased to over 95% when these features were combined with the naive Bayes classifier. If additional classifiers are used, this score has the potential to improve even further.

The co-occurrence distribution of words was established by a co-occurrence embedding proposed by **Wartena et al. [100]**. This embedding was developed to

measure the semantic links between the phrases. This embedding has outstanding performance in precision, recall, and the F1 score. The percentages of improved accuracy, recall, and F1 scores for the top 5 and top 10 terms in the ACM dataset are as follows: 9.2 and 6.9%, 10 and 8%, and 9.7 and 7.3%, respectively. These enhancements could be taken to an even higher level to achieve a precision and recall significantly higher than currently available. The method that is known as the "Co-Occurrence matrix" is the one that, according to a comprehensive analysis of the standard word embedding methods, produces results that are of a quality that is superior to those made by the other methods because it maintains the semantic connection that already exists between the words.

Word co-occurrence embedding was the method suggested by **Matsuo et al. [101]** for extracting the keyword from a single page. Correlations were created between standard terms and each term. This illustration highlighted how significant a term was inside the text. The suggested technique has a precision of about 0.5 and was later used for domain-independent keyword extraction.

A brand-new feature weighting technique called TF-IDF was put forth by **Kuang et al. [102]**. TD-IDF could not capture the significance of categorical and degree differences. The feature weighting approach was used to resolve this issue, and a significant macro F1 value of about 92.806 was obtained.

Wang [103] proposed feature extraction using weights determined by word co-occurrence. The degree of association between each text and the others was used to compute the co-occurrence rate. While classification accuracy was high, the upgraded technique appeared to be superior, with a 74% accuracy rate, and it might be improved even further by taking modified terms and synonyms into account.

The co-occurrence embedding developed by **Albathan et al. [104]** extracted high-quality patterns from the text using weighted patterns. Applying a pattern co-occurrence matrix made it possible to exclude noisy and closed sequential patterns from further consideration. Consequently, 89.04 patterns were unearthed, which allowed for an even more significant reduction in the overall quantity of noise.

Lott [105] examined the TF-IDF embedding, to find essential phrases in the document, TF-IDF embedding and naive bayes classifier could be employed. This might be expanded to incorporate co-occurrence metrics, lexical analysis, or semantic analysis.

Kadhim et al. [106] created a TF-IDF and co-occurrence embedding utilising cosine similarity to extract features. They used TF-IDF and co-occurrence embedding for this. The number of frequent words and word frequency on the page is put between the two vectors while calculating the cosine similarity score function. This creates the cosine similarity score function. Compared to TF-IDF Global, features might be modified to reduce feature space's high dimensionality and lowered by 10% to 20%. The comparison revealed both findings.

Author	Word Embedding Technique/DataSet/ Models	Model Accuracy
El-Din [94]	Bags of words, sentiment Polarity	83.5%
Dadgar et al. [95]	BBC and 20 newsgroup datasets, TF-IDF	SVM classifier=97 and 94%
Qu et al. [96]	TF-IDF	F1 score of 80–90%
Tripathy et al. [97]	SVM ,TF-IDF,Naive Bayes	94%
Enrquez et al. [98]	Word2Vec, bag of words	Accuracy of the combined BoW and Word2Vec is 75%
Soumya et al. [99]	Co-occurrence features from Wikipedia pages, bag of words	Naive Bayes classifier,F1 score =95%
Wartena et al. [100]	ACM dataset , co-occurrence embedding	9.2 and 6.9% precision, 10 and 8% recall, and 9.7 and 7.3% F1 scores
Matsuo et al. [101]	Word co-occurrence embedding	Precision of about 0.5
Kuang et al. [102]	TF-IDF	F1 value of about 92.806
Wang [103]	Word co-occurrence	74% accuracy
Albathan et al. [104]	Word co-occurrence	89.04 patterns were recovered, allowing for further noise reduction.

Table 2.1: Traditional Word Embedding

2.5.2 Static Word Embedding

The prediction-based method known as static word embedding works by transforming each word into a vector and then ascribing a probability to each word. Words are converted into dense vectors using lookup tables, and static embeddings gain new information by improving their lookup tables. This embedding is static

since it employs the same embedding tables throughout the phrase and does not change context once learned.

Zhou et al. [107] recommended utilising the Fast Text embedding of words technique in conjunction with the KNN classification algorithm, which produced an exceptional performance and was maintained in check by pruning. This combination produced excellent results. This strategy was successful for apps that categorised text that was found online. When applied to the Reuters version 3 corpus, the TC-Apte algorithm yielded pruning ratios of roughly 82.8, while the TC-PARC algorithm produced pruning ratios of approximately 72.1%.

Tezgider et al. [108] modified the settings for Word2Vec such that they were appropriate for the Turkish language, which improved word representations. The size of the vector, the number of words, and the size of the window were all measured with the assistance of some factors. We increased the quality of the word embedding by carefully considering the parameters that should be used. With a deep learning classifier model, dynamic embeddings improved the classification accuracy by roughly 89%. These embeddings also have the potential to be applied to the written text of other languages.

A word embedding called Word2Vec was introduced by **Ge et al. [109]** to do an accurate semantic similarity analysis. A graph search algorithm that could similar group traits was required for feature reduction. Comparisons were made between the Multinomial Naive Bayes, SVM, K Nearest Neighbour, and random Forest classifiers. The Word2Vec embedding technique with the SVM classifier was modified to make sense of the loose clustering method and a better-developed text corpus to achieve the highest accuracy (93,93%) possible. Both of these modifications were successful.

Mikolov et al. [110] attention was drawn to the significance of word representations that have been pre-trained and forecasted from massive text corpora like Wikipedia, Web Crawl, and news databases. The aforementioned corpus was used with pre-training tasks for Fast Text and Glove embeddings on text categorization. With an average accuracy of 82.7%, Rapid Text embedding produced the best classification performance, subsequently improved to employ this model in different tasks.

An improved word vector (IWV) word embedding approach was proposed by **Rezaeinia et al. [111]**. This method included the Word2Vec/Glove, POS tagging, and lexicon-based methodologies. This strategy was conceived to enhance the precision of word embedding vectors and developed accordingly. Afterwards, word embeddings that had previously been pre-trained were trained on a vast text corpus to produce more accurate results. The testing findings indicated that the IWV methodology had an accuracy level of roughly 80%.

Alrashdi et al. [112] presented a method for analysing tweets from the social networking site Twitter. Preprocessing the tweets involved removing any stop words, emojis, punctuation, and hashtags that might have been present. embedding in a crisis, a glove, and CNN and Bi-LSTM are different designs. With a score of 62.04% on the F1 test, the Bi-LSTM and glove word embedding method demonstrated the best performance. After that, the tweets might be further categorised using N-gram and CNN to respond to the issue.

A novel strategy developed by Facebook called Fast Text word embedding was suggested by **Kuyumcu et al. [113]**. Fast Text embedding correctly identified the text with a classification rate of 94% and has the potential to be improved for texts in other languages.

Pennington et al. [114] originated the concept of "Glove", a technique for word embedding. This model was built using the count data. The glove model successfully performed the named entity recognition, phrase analogy, and similarity tests. Using an unsupervised model-based approach, the glove generated vectors with precision and an F1 score exceeding 90%. Furthermore, there is a possibility for enhancing the glove's capabilities to undertake supplementary tasks.

Vora et al. [115] introduced a word embedding strategy for utilizing the random forest classifier to identify tweets based on emotions. Ratings were assigned to the postings made by members of Twitter depending on how effectively they communicated their feelings. The word representations for the text were developed with the assistance of the computer programs Word2Vec, Glove, and Fast Text. By categorising the text into more detailed groupings than they had been doing in the past, both these word embeddings and the random forest classifier enhanced their accuracy to 92%. Because of this, their total score went up to 94%.

Joulin et al. [116] devised a word embedding technique known as "Fast Text" and evaluated its performance on two different tasks: sentiment evaluation and label prediction. Almost one billion words were used to teach fast text in a short amount of time. Within a minute, this technique classified half a million sentences. Fast Text maximum test accuracy, which was 98.1 and 98.6% when examined with and without bigrams, is still subject to improvement.

Lilleberg et al.[117] proposed combining Word2Vec and TF-IDF word embeddings. This method uses the variations of the two approaches. This continuous word combination identified text with 90% accuracy, outperforming each individual. Each of the terms in the combination did not contain any stop words. This was achieved in conjunction with the SVM classifier. This might be improved even further by using TF-IDF and Word2Vec alongside the definitions of the words.

When **Stein et al. [118]** looked at word embeddings in conjunction with the classifiers, the results for automatically classifying documents improved. Many different word embeddings, including Glove, Word2Vec, and Fast Text, were evaluated side by side and compared. When applying the Fast Text system in conjunction with a classifier, the trials revealed that an overall rating of 0.893 was sufficient to pass the LCA F1 exam.

The PCA-based dimensionality reduction and subsequent processing methods were initially combined with the process developed by **Raunak et al.[119]**. The word vectors learned through training might be improved using this method. It is possible that decreasing the size of word embeddings will enhance the performance of devices with limited memory. The contextualised embeddings were utilised to explore further using the word embeddings influenced by the reduced embeddings, which had a maximum accuracy of approximately 90%. This resulted in a higher degree of precision.

Elsaadawy et al. [120] presented two new approaches, which they referred to as the Fast Text and Word2Vec embeddings. These techniques produce a vector by utilising a weighted average of the representation of the words as their starting point. Both models could be expanded to incorporate ensemble classifiers and offered a maximum classification accuracy of 89.93% and 90.26%, respectively, when expanded to include ensemble classifiers. Both models were also expandable.

Author	Word Embedding Technique/ DataSet/Models	Model Accuracy
Zhou et al. [107]	Fast Text, KNN classifier	82.8(TC-Apte) and 72.1%(TC-
Tezgider et al.	Fast Text, Glove, Word2Vec, deep	classification accuracy by about
Ge et al. [109]	Word2Vec, ML Models	SVM classifier had the highest accuracy (93.93%)
Mikolov et al[110]	Fast Text and Glove embeddings , text corpora like Wikipedia, Web Crawl, and news databases	Average accuracy of 82.7%,
Rezaeinia et al. [111]	Improved word vector (IWV) (Word2Vec/Glove, POS tagging, and lexicon-based methodologies)	Accuracy level of about 80%
Alrashdi et al.	Glove embedding, CNN and Bi-LSTM	F1 score of 62.04%.
Pennington et al. [114]	Glove	F1 score of about 90%
Vora et al. [115]	Word2Vec, Glove, and Fast Text, Random forest classifier	91% accuracy
Joulin et al. [116]	Fast Text	98.1 and 98.6% when examined with and without bigrams
Lilleberg et al.	TF-IDF and Word2Vec	90% accuracy in text classification
Stein et al. [118]	Glove, Word2Vec, and Fast Text	F1 score of 0.893
Raunak et al[119]	Principal component analysis (PCA), contextualised embeddings	Accuracy of about 90%
Elsaadawy et al. [120]	Fast Text, Nave Bayes, SVM,Word2Vec	Accuracy of 90.26 and 89.93%

Table 2.2: Static Word Embedding

2.5.3 Contextualized Word Embedding

This will provide terms that are similar but separate representations of context depending on the context. These representations are subject to constant change, and how they shift can be contingent on the setting in which a particular word appears.

Ethayarajh [121] investigated the contextual relevance of contextualised word embeddings like ELMo, BERT, and GPT-2. Specifically, he focused on how contextual these embeddings were. The upper levels of these embeddings generated

more context-specific representations than the lower layers. It is possible that static embedding contributes to less than five per cent of the average variation in contextualized representations.

Sarzynska et al. [122] introduced deep context-sensitive word representations and embeddings derived from language models. ELMo (Embeddings from Language Models) is a deeply contextualized word representation model. A large text corpus was utilized to pre-train the ELMo embedding. The information and words were conveyed through various context-specific grammatical and meaning-related elements. Almost 97% classification accuracy was attained using ELMo embedding and the neural network.

The BERT for labelled sentences methodology is a brand-new data augmentation approach that was proposed by **Wu et al. [123]**. When compared to other types of bidirectional models, such as shallow bidirectional models and uni-directional models, BERT, which is a deep bidirectional model, is the type of bidirectional model that is the most successful. The mask language framework that included conditional restrictions was improved, and the resulting dependent masked language model was produced as a direct consequence of these changes. The model performed far better than other embeddings, with an accuracy of approximately 79.50%.

Alsentzer et al. [124] conducted a study on clinical text utilizing the BERT model. Clinical researchers derived advantages from the utilization of these embeddings. Some models, such as BioBERT, undergo refinement, whereas others are explicitly trained in clinical literature. The biomedical research articles corpus was acquired from PubMed and utilized to train the BERT model in BioBERT. This BERT model, which has received clinical training, has demonstrated exceptional accuracy and an F1 score exceeding 90%. Furthermore, it can be effectively utilized in many clinical NLP research endeavours.

Chang et al. [125] were the original proponents of the concept now known as X-BERT. This strategy will optimize the BERT embedding to achieve optimal performance. The issue of tracking contacts in BERT that occurred in the past has been rectified. This specific embedding attained a precision rate of 68%, and many fine-tuned BERT models further enhanced it to reach outstanding performance.

Jwa et al.[126] The "BAKE" model for detecting fake news uses BERT embedding. "exBAKE" refers to BAKE-contributed unlabelled news. Algorithms determined phoney news in these models. Because they analyzed headlines and news item body text, these models worked with the FNC-1 dataset. They scored 0.125 and 0.137 points higher than their competitors in F1. Adding news during pre-training could improve this.

Maslennikova [127] tested Word2Vec embedding with ELMo. ELMo embedding improved binary classification and prediction accuracy. Using ELMo embedding resulted in a 10% improvement in the F1 score compared to Word2Vec.

Chakraborty et al.[128] examined and analyzed a sparse vectorizer. It performed far better than neural word embeddings such as Word2Vec, Glove, Fast Text and character embeddings such as ELMo. This result was reached after analyzing the effectiveness of the sparse vectorizer and earlier embeddings. The execution was enhanced thanks to this embedding, which led to a 3–5% increase in the F1 score.

For the visually impaired, **Manoharan [129]** devised a simple mechanism for image processing that recognises the text and vocalisation. The images from the scanner were processed using the onboard LattePanda Alpha system. Images were sorted into similar alphanumeric characters after being pre-processed, segmented, and extracted of characteristics. The voice synthesiser was used to turn the text data into audio output. This model provided an accuracy of approximately 97%, and it was improved even further by employing a machine learning method to analyse textual data, which increased the system's accuracy and reduced the processing time.

Schwartz et al. [130] investigated the possibility of using a mixture of two-word level embeddings to carry out binary classification tasks. The combination of ELMo and BERT could have improved the categorisation ability. ELMo or BERT, when paired with the Glove embedding, can increase classification performance, with a maximum x2 value of between 12 and 16.69. This performance can be improved by analysing the algorithms' effectiveness on several datasets.

Domain adaptive fine-tuning is an easy method proposed by Büyükoz et al. [131] for applying unsupervised labelling to brand-new domains. The process in question is called domain adaptive fine-tuning. Using masked language modelling within the

text, the contextualised embeddings were altered. The ELMo and BERT embeddings also significantly improved due to some fine-tuning, which produced excellent 83 %results.

Wang et al. [132] carried out a series of controlled tests to systematically investigate traditional word embeddings in addition to contextualised versions of these embeddings for the aim of text classification. According to the findings, BERT performs significantly better than ELMo in general, particularly for long document collections.

Among the various models, **Joni Salmine et al. [133]** offered the fakeRoBERTa model based on GPT2 as having the highest-performing accuracy of 96.64%.

Author	Word Embedding Technique/DataSet/ Models	Model Accuracy
Ethayarajh [121]	ELMo, BERT, and GPT-2	
Sarzynska et al. [122]	ELMo embedding	97% classification accuracy
Wu et al. [123]	BERT	Accuracy of about 79.60%
Alsentzer et al. [124]	BERT	F1 score above 90%
Chang et al. [125]	X-BERT	Precision rate of 68 percent
Jwa et al. [126]	BERT, FNC-1 dataset	0.125 and 0.137 F1 scores
Maslennikova [127]	ELMo	F1 score was 10% higher than Word2Vec embedding.
Chakraborty et al. [128]	ELMo	Enhanced F1 score by between 3 and 5%
Manoharan [129]	Image processing,	Accuracy of about 97%
Schwartz et al. [130]	Combination of ELMo and BERT	Accuracy of about 83%
Wang et al. [132]	ELMo, BERT	BERT performs significantly better than ELMo
Joni Salmine et al. [133]	fakeRoBERTa, GPT2	fakeRoBERTa predict accuracy of 96.64%.

Table 2.3: Contextualized Word Embedding

Using contextualised word embeddings has succeeded in the most critical NLP tasks. Besides this fact, many aspects of the topic still need to be investigated. This part gives some challenging problems and potential future paths for research.

Worsham et al. [134] utilized a multitask deep neural network approach for pre-training the BERT and learning the MTL language representation. Because of this, the network achieved the highest possible score on the GLUE benchmark, 82.2%.

The term "shot learning" refers to training a learning model with limited data.

This is referred to as "learning with a shot." This practice is common in machine learning. Components of traditional treatments are often hand-crafted by the practitioner. Its primary focus is on metric learning, a method founded on the distance metric that looks for similarities and differences between various items. In other words, it compares and contrasts them. Consequently, embedding embeddings focused on providing single-shot learning examples will be necessary. According to the findings of **Hou et al. [135]**, the most successful strategy to eliminate representation ambiguity is using few-shot learning in conjunction with pairwise embedding and a unique query support arrangement.

ELMo, BERT, and GPT pre-training models use self-supervision. Stitching an unlabelled corpus together to generate tagged data depends on sentence and word co-occurrence. Despite its importance, syntactic, semantic, and lexical information has yet to be addressed. **Sun et al. [136]** recommend a continuous pre-training approach for MTL task learning. Semantics, structures, and word levels are studied in many pre-training activities, yielding fruitful results. So, this paradigm is necessary to learn stable representations by carefully arranging challenging activities.

The following is a list of some of the fake review models that have been built over the past ten years, together with information regarding their correctness, the dataset source, the name of the research paper, and the false review parameters:

DeepFakeDet- Deep learning is the methodology applied by Wang et al. [157]'s model, which was built to identify and counteract the effects of fraudulent reviews. The system was trained using a dataset consisting of 200,000 customer reviews from Amazon, and it obtained an accuracy of 90%.

ReviewTrust- This model to identify fake reviews was developed by Liu et al. [158], and it combines a combination of text features, behavioural variables, and social network information. It attained an accuracy of 92% after being trained on a dataset consisting of one million customer reviews from Amazon.

FakingIt- This algorithm was developed by Verma et al. [159], using a method known as deep learning to identify false reviews. It was trained using a dataset consisting of 10 million customer reviews from Amazon and obtained an accuracy of 93%.

FakeReviewerDet- This model to identify fraudulent reviewers was developed by Zhang et al. (2020), and it uses a combination of text features, behavioural factors, and social network information. It obtained an accuracy of 94% after being trained on a dataset consisting of 10 million reviews from Amazon.

Model name & Year	Database	Accuracy	Fake review parameters	Artificial models	Authors
FakeReview Net & 2019	Amazon product reviews	88.20%	Word count, sentiment, helpfulness votes	Deep neural network	Zhang et al. [146]
FakeReview GAN & 2019	TripAdvisor hotel reviews	92.10%	Word count, sentiment, grammar	Generative adversarial network	Li et al. [147]
FakeReview Forest & 2019	Amazon product reviews	87.10%	Word count, sentiment, helpfulness votes	Random forest	Wang et al. [148]
FakeReview SVM & 2018	Amazon product reviews	85.70%	Word count, sentiment, helpfulness votes	Support vector machine	Chen et al[149]
FakeReview LSTM & 2017	Amazon product reviews	86.40%	Word count, sentiment, helpfulness votes	Long short-term memory neural network	Zhang et al. [150]
FakeReview BERT & 2022	Amazon product reviews	90.20%	Word count, sentiment, helpfulness votes, named entities, sentiment lexicons	BERT	Yang et al.[151]
FakeReview GAN-BERT & 2021	Amazon product reviews	91.30%	Word count, sentiment, helpfulness votes, named entities, sentiment lexicons	BERT + GAN	Chen et al.[152]
FakeReview Graph & 2020	Amazon product reviews	92.20%	Word count, sentiment, helpfulness votes, social network relationships	Graph neural network	Zhang et al. [153]

Model name & Year	Database	Accuracy	Fake review parameters	Artificial models	Authors
FakeReview-T5 & 2020	Amazon product reviews	91.10%	Word count, sentiment, helpfulness votes, natural language inference	T5	Wang et al. [154]
FakeReview-RoBERTa & 2022	Amazon product reviews	91.20%	Word count, sentiment, helpfulness votes, contextual word representations	RoBERTa	Zhang et al. [155]
FakeReview-DPR & 2022	Amazon product reviews	91.50%	Word count, sentiment, helpfulness votes, document-level representations	DPR	Zhang et al. [156]
fakeRoBERTa & 2022	Amazon product reviews-OSF	96.64%.	Context Level	GPT, RoBERTa	Joni Salmine et al. [133]

Table 2.4: Fake Reviews Models

2.6 Research Gap

The following are some research gaps regarding the identification of fake reviews using classic word embedding techniques:

The **semantic meaning of words** cannot always be accurately captured by employing word-embedding techniques. This is because word embeddings are often trained on vast text corpora, which may or may not contain all the words utilized in fake reviews.

Word-embedding approaches are only sometimes reliable when accurately capturing the **context of individual words**. This is because word embeddings are often trained using bag-of-words models, which do not consider the relationships between phrases. The traditional word embedding methods do not think the surrounding text. Because of this, they may be unable to differentiate between words whose meanings are comparable but which are used in different situations. For instance, "great" can convey either a good or a negative attitude, depending on the context.

The conventional word embedding methods cannot always recognise the **patterns** utilized in fake reviews. Specific patterns may be complex to discern due to their complexity. For instance, deceptive reviews frequently use exaggeration or language that conveys emotion.

2.7 Research directions explored to address the gap

By utilising **neural networks or domain-specific corpora**, it may be possible to build new word embedding approaches that are more effective in capturing the semantic meaning of words.

Employing dependency parsing or semantic role labelling can generate **novel word embedding methodologies** that are more proficient in obtaining the contextual meaning of words. These techniques can be utilized for the development of novel word embedding methods.

Ensemble learning can be used to develop new methods for combining word embedding techniques with other features, such as sentiment analysis or product features. These methods can also create new approaches for mixing word embedding techniques with additional features.

Key elements that reflect Methodological Quality Assessment for literature review may include data collection and preparation encompassing the suitability and variety of the datasets employed for training and testing deep learning models. This entails verifying the precision of labelling (differentiating authentic from fake reviews) and ensuring that the datasets cover a diverse range of review categories and potential patterns of fake reviews. The study assesses the level of complexity and suitability of the deep learning structures employed in the research, encompassing their model architecture and training techniques. Furthermore, evaluating the training methodologies and hyper-parameter optimization procedures is crucial to avoid overfitting or underfitting the data by the models. When identifying fake reviews, it is essential to incorporate performance measurements. However, it is necessary to acknowledge and express the limitations and potential biases associated with these metrics.

The domain of fake review identification utilising machine and deep learning models are continuously progressing, with scholarly investigations showcasing both encouraging outcomes and contradictory discoveries. Certain studies demonstrate high levels of accuracy in their models. In contrast, others emphasise the challenges in addressing particular fraudulent reviews or developing new strategies malicious individuals adopt. These inconsistencies can arise from various factors, such as data

diversity, model complexity, and the selection of evaluation metrics. Data diversity refers to the variations in the datasets utilised for training and testing machine learning models. These variations can substantially impact the outcomes and results obtained from the models. Research using datasets significantly biased towards a specific sort of fraudulent review may face difficulties in generalising and identifying other subtle forms of false material. Model complexity refers to selecting a model's structure and complexity, which can impact its performance. More complex models may provide quicker training, but they may need help dealing with more nuanced patterns found in fraudulent reviews. On the other hand, excessively intricate models might result in overfitting, which impairs their capacity to be applied to new data.

The selection of evaluation metrics can also impact the perceived efficacy of models. Metrics such as accuracy can be deceptive since they may need to consider the capability to identify fraudulent reviews if the focus is only on correctly categorising genuine reviews. An all-encompassing approach considers precision, recall, and F1 score measurements to offer a more nuanced perspective.

Chapter 3

Research Methodology/Proposed Framework

The model, named FRARBiLSTM (Fake Reviews-AFINN Roberta using Bidirectional LSTM), has been dissected into its components, as illustrated in Figure 3.1. The initial stage encompasses both the collection and preparation of data. The data undergo preprocessing through Natural Language Processing (NLP) techniques, including removing stop words and punctuation, converting to lowercase English characters, and stemming. Additional NLP operations encompass the following. The second phase involves embedding data and enhancing text with sentiment polarity, which ensures the dataset's good quality during the feature selection process. This will require the utilisation of Word embedding methodologies, such as Roberta, to convert texts into numerical representations. The last stage of our proposed model involves the identification of spam. During this stage, traditional machine learning and deep learning classifiers classify reviews as either spam or legitimate (ham).

FRARBiLSTM model built using AFINN, RoBERTa(POS+LIWC), and Bidirectional LSTM approach for Novelty purposes that generates the best result when compared with the existing model that worked on the same benched dataset.

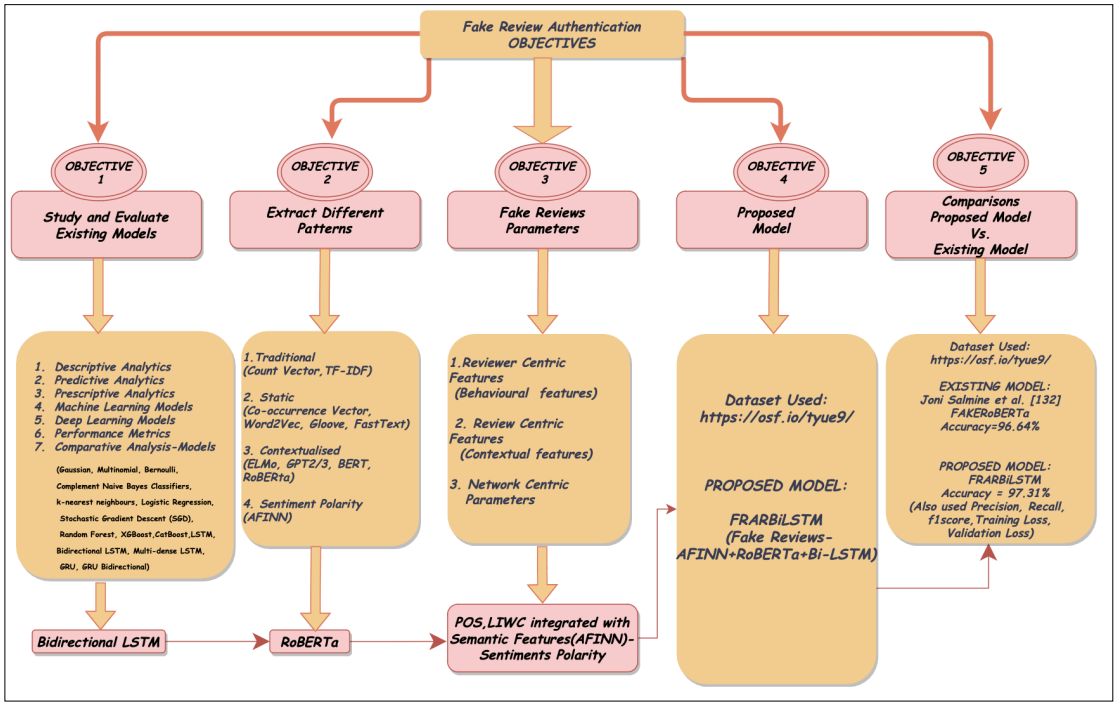


Figure 3.1: The Proposed Framework

3.1 FRARBiLSTM Model-Criterion to Design-Objective wise

As per Objective 1

Firstly, we focus on descriptive analytics in which we focus on the dataset that we considered from the genuine source:https://osf.io/tyue9/.Then we implement exploratory data analysis on the dataset to remove the anomalies from the dataset. Subsequently, we apply 10 conventional machine learning and five deep learning models to the primary dataset to assess the efficacy of the most optimal model. So here, the Bi-LSTM neural network performs best among all models, with the implementation part shown in the result section.

As per Objective 2

Under objective 2, we have extracted different patterns from the reviews that help the models for feature engineering. We implement **TF-IDF** on our traditional machine learning models and **RoBERTa**, one of the latest context-based data extraction techniques. In this phase, we also extract sentiments of reviews using **AFINN**, which also be considered one of the fake review parameters that help in feature engineering.

As per Objective 3

In this phase, we study 1**4 reviewer-specific, 7 review-specific, and 5 network-specific parameters**. However, per my database, we only review parameters such as **POS, LIWC, and Semantic features**. We ensemble these three features with RoBERTa, AFINN, and BiLSTM to help constitute our new model under objective 4.

As per Objective 4

Under objective 4, we proposed our new **FRARBiLSTM model**, which we implemented on the same dataset where we implemented our 10 traditional ML and five DL models. In the implementation phase, we also use NLP techniques like sentence tokenization, stemming / lemmatization, stop-word removal, etc.

As per Objective 5

In this final phase, we **compare our FRARBiLSTM model** with the existing one using the same database source mentioned under objective 1. Comparisons based on

parameters like accuracy, precision, recall, f1score, POS, LIWC, Accuracy, and loss graphs will finally be shown in the results section.

3.2 Preprocessing

In Phase I of our suggested paradigm, we will examine the Acquiring and the Preprocessing of Data.

Data Acquisition

A limited number of datasets comprise authentic assessments of superior quality and reviews deliberately manipulated to deceive readers. To begin, compile an extensive database containing both authentic and counterfeit evaluations. The dataset should comprise equivalent genuine and fabricated reviews to ensure balance. Multiple websites offer the dataset for download, including OSF, Amazon, Yelp, and TripAdvisor. For this investigation, the OSF Fake Review Dataset, accessible at https://osf.io/tyue9/, was utilized.

Data Preprocessing

During the preprocessing phase of the project, tokenization was used to split the text into individual words. Lowercase letters were also used, and punctuation marks such as periods and colons were removed. Data preprocessing is often considered one of the most critical stages of any machine learning technique. The preprocessing of data is required since the data that exists in the world is never in a usable form. In this investigation, various preprocessing procedures were carried out to prepare the raw data from the dataset for subsequent analysis. In the following, an explanation of the preprocessing methods used in the suggested framework is provided

a) **Tokenization-** The individual words that comprise the text are referred to as tokens. For example, tokenization will break the line "I love the look and feel of this pillow" down into its parts, which are referred to as tokens: "I," "love," "the," "look," "and," "feel," "of," "this," “pillow."

b) **Eliminating Stop Words-** Stop words are the terms that are used the most frequently [24], even though they have no real significance. The words "an," "a," "the," and "this" are all examples of stop words. Before continuing with the approach in this study that detects fake reviews, all of the data are cleaned of stop words.

c) **Eliminating All of the Punctuation-** The text is broken up into sentences, paragraphs, and phrases with the help of punctuation. Punctuation marks affect the results of any technique of text processing because they are used so frequently in written communication. This is especially true of methods predicated on the frequency with which individual words and phrases are employed.

d) **Lowercasing-** The transformation of letters written in uppercase into letters written in lowercase was the only pre-processing strategy that significantly outperformed the baseline test result. Although words like "Book" and "book" signify the same thing, the models interpret them differently depending on whether or not they are written in lowercase letters.

e) **Stemming-** In English, a single phrase can have several meanings depending on how it is phrased. These variances in a source text led to redundant data when it came time to create NLP or machine learning models. These models won't operate very well. It is necessary to standardise the text by removing any instances of repetition and reducing terms to their root forms. This is a prerequisite for this process.

f) **Eliminating Frequent and Rare Words-** Assuming that the counts of the dataset's frequent words are pretty high, most scoring systems award points for correctly identifying those words more than they award points for correctly identifying the counts of other words. This gives the impression that every other word occurs less frequently. A different justification leads to the elimination of uncommon terms. The noise caused by their rarity makes it difficult to make meaningful connections between them and other words.

3.3 Data Enrich Text

We suggested employing the AFINN lexicon to ascertain the sentiment polarity to enhance text columns. An alternative is to utilise a computer programme that relies on a dictionary, such as the AFINN model. Each word in the dictionary can be associated with a particular emotion. When we tokenise a statement, we assign a score to each word to signify its positive or negative contribution to the overall meaning of the statement. The last phase entails assigning a comprehensive rating to

the statement. The computation result will determine if positive or negative words predominantly characterize a statement.

3.4 Feature Selection

The feature selection is the subsequent phase of our suggested model. The TF-IDF and RoBERTa are utilized to develop our mythology. Finding and collecting necessary information from the review data is known as **feature extraction**, so the model may always be trained. It is a significant step in developing a deep-learning model for identifying fake reviews.

The performance of either a pattern recognition system or a machine learning system can be improved by using feature extraction, which is the primary goal of this process. Feature extraction is the method that decreases an input by removing its fundamental features to give more valuable data for models.

The most crucial stage is to clear the data of any extra features, which can have the unintended consequence of lowering the model's accuracy. Within this field of research, several academics have examined various techniques for extracting features. These techniques encompass sentiment analysis, polarity, text length, word frequency, and n-grams. The study work utilises TfidfTransformer and TfidfVectorizer, which are essential NLP tools, to extract features from text input and represent those features. They are a constituent of the sci-kit-learn library, which can be located in Python.

Conversely, the TfidfVectorizer integrates the processes of extracting features and representing features into a unified function. This enhances efficiency. The input for the process is the raw text data, and the output is in the form of Tf-Idf vectors. TfidfVectorizer outperforms TfidfTransformer in handling the earliest stages of text pre-processing and cleaning, which include eliminating stop words, stemming and lemmatizing, and transforming the text into a numerical format. Moreover, TfidfVectorizer converts the text into a vectorized format, which can be used in other applications. In addition, TfidfVectorizer can transform text into a vectorized representation.

The Tfidf Transformer converts raw count-based feature vectors into their equivalent Tf-Idf representations. The Tf-Idf score of a word is determined by considering its

frequency within a particular document and its scarcity throughout the entire corpus. The Tfidf Transformer and Tfidf Vectorizer are widely used in text clustering and sentiment analysis applications.

Tokens with a high frequency in a document corpus are assigned lower importance in the analysis than tokens that appear in a significantly smaller portion of the training corpus. This is because the training corpus contains a more significant number of documents. This is because the TF-IDF metric rescales features instead of relying solely on the raw frequency of token occurrences in a document corpus.

3.5 Fake Review Detection

We employed five deep-learning algorithms, namely LSTM, Bidirectional LSTM, multi-dense LSTM, GRU, and Bidirectional GRU, to detect false reviews using hybrid and Ensemble modelling techniques. The collection also includes ten conventional machine learning classifiers: Gaussian, Multinomial, Bernoulli, Complement NB Classifiers, KNN, SGD, Logistic Regression, RF, CatBoost, and XGBoost.The first step of this process is to detect fraudulent reviews by examining the results collected from eight machine-learning models. Finally, the penultimate stage of the process involves choosing an appropriate deep-learning model to identify fraudulent reviews. This study utilised five deep learning models to analyse the same dataset, including LSTM, Bi-LSTM, Multi-dense LSTM, GRU, and Bidirectional GRU. This was undertaken to achieve the goals of this project. Data are utilised in feature extraction, training, and testing, respectively.

A model is trained, and its performance is evaluated using this data. The design is constructed by adjusting its variables based on the training data to decrease errors from making predictions using the same training data. This is accomplished by utilising the training data. At this stage, the efficiency of the training strategy is assessed by analysing the data collected from the different tests. This assessment is being conducted using the collected information. The model is estimated using this dataset to determine its ability to generalise to unseen data and accurately predict future experimental outcomes. It is usual to keep the testing data distinct from the training data. This is done to prevent the model from becoming overfitting. Part of the procedure involves dividing the dataset into two separate subsets. The prediction model is first adapted to the training dataset, the first subset in this process. This step

tests the machine learning model's performance on untrained data. New data is model-untrained data.

We will implement the model in this manner. It is necessary to align the information with established data, including input and output values, to provide greater clarity. Additionally, we must consider future scenarios when the desired outcomes or expected results may not be available. The train-test strategy is suitable when there is a dataset of a manageable size.

3.6 Classification Models Used

A. **Naive Bayes (NB)-** The Bayes theorem forms the foundation of the NB model's central idea. NB arrives at a set of probabilities by calculating the frequency of occurrence in a dataset and then tallying its total value. NB has been utilized successfully in various application sectors, including text classification, spam filtering, and recommendation systems. The Naive Bayes method is applied to find solutions to problems associated with classification. Bayes's theorem forms the foundation for this technique. Text classification, which frequently uses high-dimensional training datasets, is this technology's primary application. The Naive Bayes approach is a classification strategy that is easy to understand and practical. It contributes to constructing rapid machine learning models that promptly provide correct predictions. It is a probabilistic classifier which bases its predictions on the possibility that an object will occur, and it uses this likelihood to make its classifications.

B. **K-Nearest Neighbours (KNN)-** KNN categorization is simple and effective. KNN works best for statistical estimates and pattern identification. KNN uses comparable cases to classify instance queries. Distance is usually used to calculate similarity. It is used to classify locations based on the assumption that nearby sites are close. Regression is less prevalent than categorization. Classification problems are solved by majority voting. In actual literary works, "majority vote" is used instead of "plurality voting," which is used in actual voting. "Majority voting" requires a majority above 50% and works best when voters can choose between two groupings.

C. **Decision Tree-** This is yet another machine learning classifier that focuses on building a tree to represent a decision made based on the training data. The method constructs the tree iteratively, beginning with the optimal feature split as the starting point. A function that has been determined in advance is utilised to select the finest qualities. This function could be entropy, information gain, gain ratio, or the Gini index. A decision When you train with Tree, it generates a model that you may employ for making predictions regarding the category or value of the target variable. This is performed by gaining knowledge of easy decision rules deduced from historical data, also known as training data.

D. **Support Vector Machines (SVM)-** SVM separates data into classes. The best separable hyper-plane correctly classifies the training data. The SVM algorithm determines a hyperplane that can classify data points in a space with N dimensions (N being the number of characteristics). Many hyperplanes can distinguish the two data points. We want a plane with the most significant margin or distance between both classes' data points. When maximising the margin distance, reinforcement helps identify the following data points. Support vectors are data points near the hyperplane that affect its position and direction. Support vectors maximise classifier margin. Eliminating support vectors moves the hyperplane. These factors aid SVM development.

E. **Random Forest-** One of the practical solutions to the overfitting problems that can develop in the decision tree is Random Forest. The primary idea behind a random forest is to construct a forest using different datasets as individual tree samples. Instead of making each tree in the forest from all available features, Random Forest chooses a small number of elements randomly before constructing the tree. It can also be used for regression analysis in addition to classification analysis. Random forests are created like decision trees by constructing decision trees. Multiple trees are formed in a manner that is parallel and independent of one another. All training examples run via substitution are incorporated into each tree while it is being constructed. The parameters at every tree node are improved after the forest of decision trees has been built. Constructing a forest of decision trees allows for improvements to be made to the parameters at each tree node. During training on independent trees, each tree node

will eventually approach a different, randomised subset of the collection of features.

F. **SGD-Stochastic Gradient Descent-** This general-purpose optimisation technique may discover the most effective answers to various challenges. The steps' size, defined by the learning rate hyper-parameters, is an essential part of Gradient Descent (GD), which stands for gradient descent. If the learning rate is too low, the algorithm will have to go through many iterations to converge, which will take a significant amount of time; on the other hand, if the learning rate is too high, we might skip over the optimal value.

G. **XGBoost-** XGBoost is a highly optimised distributed gradient boosting toolkit designed to train machine learning models efficiently and scalably. It aggregates the predictions of multiple less precise models to obtain a more accurate forecast. XGBoost, or Extreme Gradient Boosting, is gaining widespread recognition as a highly effective machine learning technique for tackling intricate classification and regression issues. XGBoost's ability to address real-world data that effectively contains missing values is one of its most essential aspects. This enables the algorithm to do so without requiring substantial data pre-processing. In addition, XGBoost features built-in support for parallel processing, making it possible to train models on big datasets quickly.

H. **CatBoost-** CatBoost is a supervised machine learning approach utilized by the Train Using AutoML tool. This method employs decision trees for classification and regression analysis. CatBoost uses gradient boosting, which is hinted at by its name, and works with categorical data, hence the name "Cat." The "gradient boosting" procedure involves the repeated construction of several decision trees. Each consecutive tree enhances the result of the tree before it, ultimately leading to more significant results. CatBoost is an improved and significantly quicker implementation of the original gradient boost algorithm. CatBoost can overcome a limitation that is present in other decision tree-based methods. The data must typically be preprocessed in these methods to transform category string variables to numerical values, one-hot-encodings, etc. CatBoost can circumvent this limitation. This technique can immediately ingest a mix of categorical and non-categorical explanatory variables without any prior preprocessing. The algorithm

incorporates it as a preprocessing step. CatBoost employs a method known as ordered encoding to encode categorical feature data. When calculating a value to replace an absolute characteristic, requested encoding considers the target statistics for all rows before the data point is encoded. CatBoost is distinguished from other similar programs in that it uses symmetric trees. This indicates that every decision node uses the identical split condition throughout all depth levels. CatBoost may also be faster in some situations than other approaches, such as XGBoost. The previous algorithms' cross-validation, regularisation, and missing value support are among the features carried over into this one. This strategy works very well with trim and a lot of data.

I. **Long-Term Short-Term Memory (LSTM)-** LSTM solves the difficulties of vanishing and exploding gradients that occur in regular RNNs. This difficulty arises when the RNN is trying to learn anything complex. This issue manifests itself when the weights of the RNN need to be updated at a faster rate. LSTM is a fundamental component of deep learning. Memory cells are responsible for this ability, enabling the model to store and update information over time. Recently, several improvements have been made in the field of LSTMs. One study released in 2021 offered a new sort of LSTM named the "coupled LSTM," which outperformed standard LSTMs on various tasks. Another work that was published in 2022 presented a revolutionary LSTM-based architecture called "GLSTM," which made use of a gated linear unit (GLU) activation function [137]. In general, LSTMs continue to be widely used since they are such an effective tool for processing sequential data, and current research continues to investigate novel design variations and ways to improve them.

J. Bidirectional LSTM (BLSTM)- BLSTM [138] is an enhanced version of the conventional LSTM architecture that handles the input sequence in both the forward and backward directions. As a result, the model can comprehend the connections between the present input and previous and forthcoming inputs. One of the LSTM layers of a BLSTM [139] processes the input sequence in the forward direction, while the other layer processes it in the reverse order. The input sequence is partitioned into two distinct LSTM layers. Once a layer is completed, the outcomes are aggregated and passed on to the subsequent layer.BLSTMs were

first introduced to the public by Schuster and Paliwal in 1997. Since then, they have been widely utilised in several applications, including language translation, speech recognition, and image captioning. The primary goal of recent research and development efforts in the field of BLSTMs has been to improve the devices' overall levels of both accuracy and efficiency. One piece of research completed in 2021 and then released the following year introduced a novel type of BLSTM called the "depth-wise bidirectional LSTM." Another study was carried out and published in 2022; it presented a ground-breaking BLSTM-based design and dubbed it “FusionLSTM." In general, BLSTMs continue to be an effective tool for processing sequential data, and continuing research continues to investigate novel versions of and improvements to the model's architecture.

K. Multi-dense LSTM Model- A Multi-dense LSTM Model is a neural network design utilized in deep learning. It employs a combination of densely coupled layers and LSTM layers. The dense layers process the incoming information and extract significant insights. On the other hand, the LSTM layers are responsible for processing sequence input and retaining the recollection of previous events to impact future predictions. The Multi-dense LSTM Model is a highly effective tool for analyzing sequence and non-sequential input. It is well-suited for tasks like sentiment analysis and time series prediction. The combination of dense and LSTM layers allows for effectively capturing connections within the data. This enables it to capture non-sequential and sequence relationships within the data efficiently.

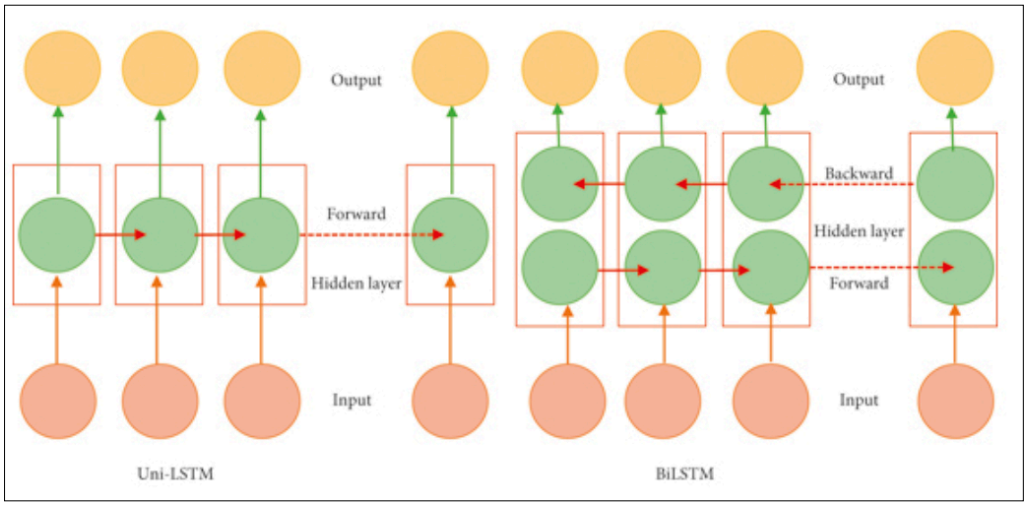


Figure 3.2: Uni-LSTM and BiLSTM Architecture[140]

L. GRU- The purpose of GRU is to handle sequential input and retain a lasting memory of past occurrences to impact future predictions. This was achieved to enhance precision. GRUs, unlike conventional LSTM and RNN systems, possess two additional gates. Because of this, GRUs can more efficiently capture the connections between items in sequential data while improving their computing efficiency and making it more straightforward to train them.

M. GRU for Bi-Directional- The Bidirectional GRU (BiGRU) in deep learning is a modified version of the GRU network that analyses sequential input in both the forward and backward directions. To comprehensively represent the data, a BiGRU network will analyse the sequence in both the forward and backward directions and subsequently combine the resulting hidden states. BiGRUs can capture contextual information and dependencies in series, which makes them successful in various sequential data processing tasks, including sentiment analysis and speech recognition. BiGRUs provide a more robust and adaptable option for handling sequential data than ordinary GRUs or other types of RNNs due to their ability to combine the advantages of GRUs with bidirectional processing. This is achieved by synergistically leveraging Recurrent Neural Networks (RNNs) and Gated Recurrent Units (GRUs) capabilities.

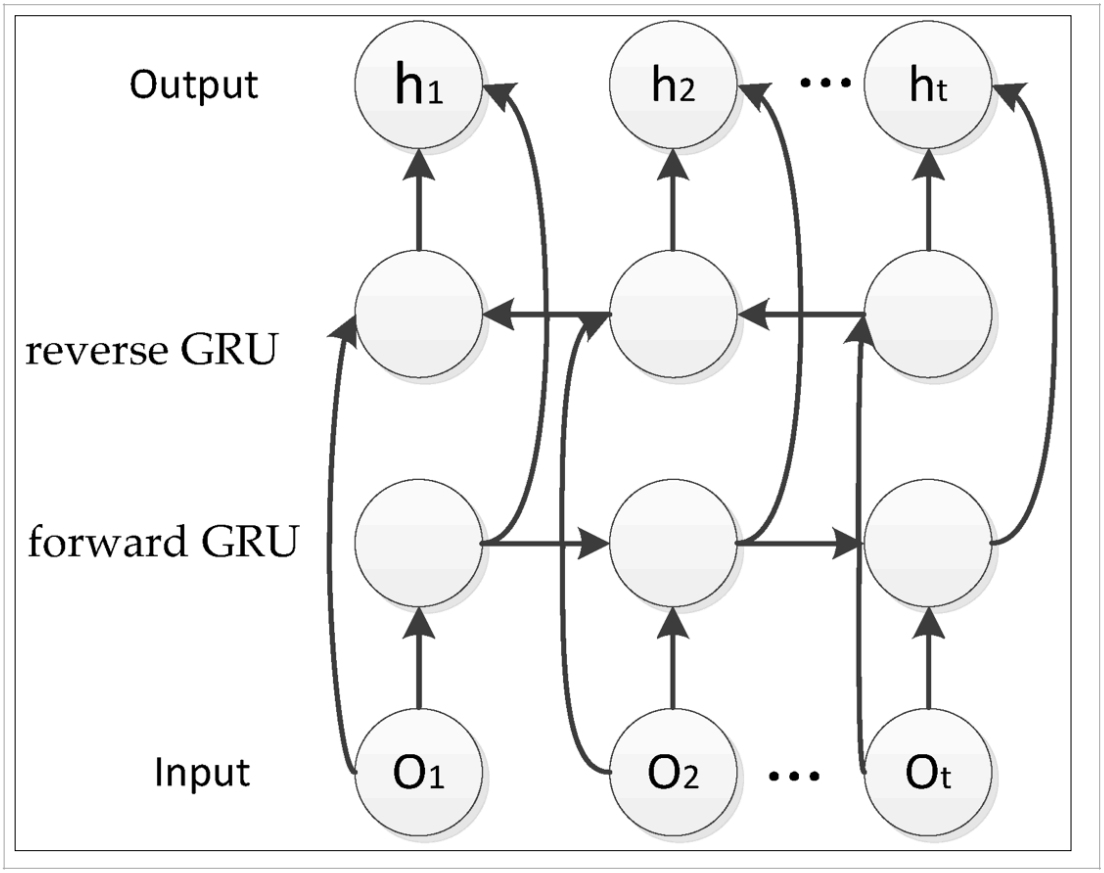


Figure 3.3: BiGRU Network Layer[141]

3.7 Ensemble Modelling

The Ensemble and Hybrid Modelling techniques are two prevalent methods used to construct more accurate models for identifying fake reviews. When multiple models' predictions are merged in ensemble modelling, a more dependable and precise model is produced due to the process. The concept underlying ensemble modelling is that by combining different models, the potential advantages of each model can be optimised. In contrast, the possible drawbacks of the models might be avoided. Ensemble learning is a collection of approaches that combine predictions from numerous models to achieve higher predictive performance, and it is one of the methods we have used here. Its tactic is known as "strength in unity," it relies on the fact that efficient combinations of weak learners can produce more accurate and robust models. The three primary categories of ensemble learning methods are known as "bagging," "boosting," and "stacking," respectively. The data presented in the table above makes it clear that boosting techniques, such as XGBoost and CatBoost, can outperform typical machine learning algorithms. Then, these boosting methods were implemented with RoBERTa to improve the result. The accuracy of

the results that XGBoost and CatBoost generate while utilising the RoBERTa Algorithm is 91% and 92%, respectively.

The following are some examples of the several types of ensemble models-

Voting Classifier- A Voting Classifier aggregates the predictions of multiple models to determine a single conclusion, which is the most frequently agreed upon by all models. The process involves conducting a vote, wherein the model that receives the highest number of votes emerges as the winner.

Bagging- Bootstrap Aggregating, or "bagging," involves creating several samples from the model training dataset and training different models on each sample.

The term "bagging" refers to this process. The aggregate prediction takes into account the findings of all of the models and employs either the models' average vote or the model with the majority vote.

Boosting- Boosting is a method for training a series of models one after the other, with each model trying to fix the mistakes made by the models that came before it during the training process. Boosting is also written as "boosting" and “boosting."

Stacking- When you stack, you train many models and then feed their results to a higher-level model. It's called "stacking up."

Ensemble modelling can help make fake review spotting algorithms work better because it can get around the problems that single models have and lower bias.

3.8 Hybrid Modelling

Artificial intelligence (AI) is the encompassing technology that encompasses a wide variety of developments and fields of application. Some include robots, natural language processing, intelligent systems, machine learning, etc. During training, an iterative process is used to identify model parameters to maximise the association between the predictor and the targeted variables. This will allow for maximum prediction accuracy. When new predictor data is given, the trained model estimates the target variable using the previously identified pattern. Figure 3.4 shows the workflow schematically.

Throughout the past four centuries, this workflow has been consistently utilized. The hybrid machine learning (HML) workflow is an innovative new strategy that improves the conventional method.

It is necessary to explain this new learning workflow in the context of comprehending the machine learning tools we use. It is essential to grasp what goes into the algorithms to understand how they operate. The more we learn how they function, the more transparent they appear, and the more we can diminish the "black box" phenomena around them.

Hybrid modelling seeks to enhance the precision of a model by integrating multiple models or data sources. A practical approach to using mixed modelling to detect fraudulent reviews is incorporating many features. For instance, features based on language, metadata features, and behaviour-related characteristics.

Hybrid modelling refers to integrating various models into a more advanced model. This can be achieved by mixing text-based and metadata-based models and utilising the outputs of both models as input features.

Hybrid modelling, by including the most valuable elements from many models and data sources, can enhance accuracy and comprehensiveness compared to other models. Mixed modelling incorporates the most relevant components from multiple models. However, the construction and maintenance of it may pose additional challenges, requiring a significant allocation of resources to account for the increased difficulty resulting from its greater complexity.

Hybrid Modelling and Ensemble Modelling are helpful techniques that can be used to improve the accuracy and robustness of algorithms used for detecting fake reviews. The problem determines the strategy selected, the already available data, the resources, and the expertise currently available for building a model. artificial intelligence is an area of computer science that focuses on designing and developing intelligent systems. These systems may be implemented in hardware, software, or both. Most of us have undoubtedly been using HML algorithms in some form or another without even being aware of it. We may have utilized methods that combine existing ones or techniques that are imported from other fields in combination with existing ones. Before handing our data to a machine learning approach, we

occasionally alter techniques like principal component analysis (PCA) or fundamental linear correlation analysis. Some practitioners employ evolutionary algorithms to optimise the parameters of pre-existing machine learning methods. An artificial intelligence architecture that is only marginally distinct from the standard workflow underpins the HML algorithmic framework. We have come to take machine learning algorithms for granted because we utilise them off the shelf, most of the time, without giving any thought to the specifics of how the various components interact.

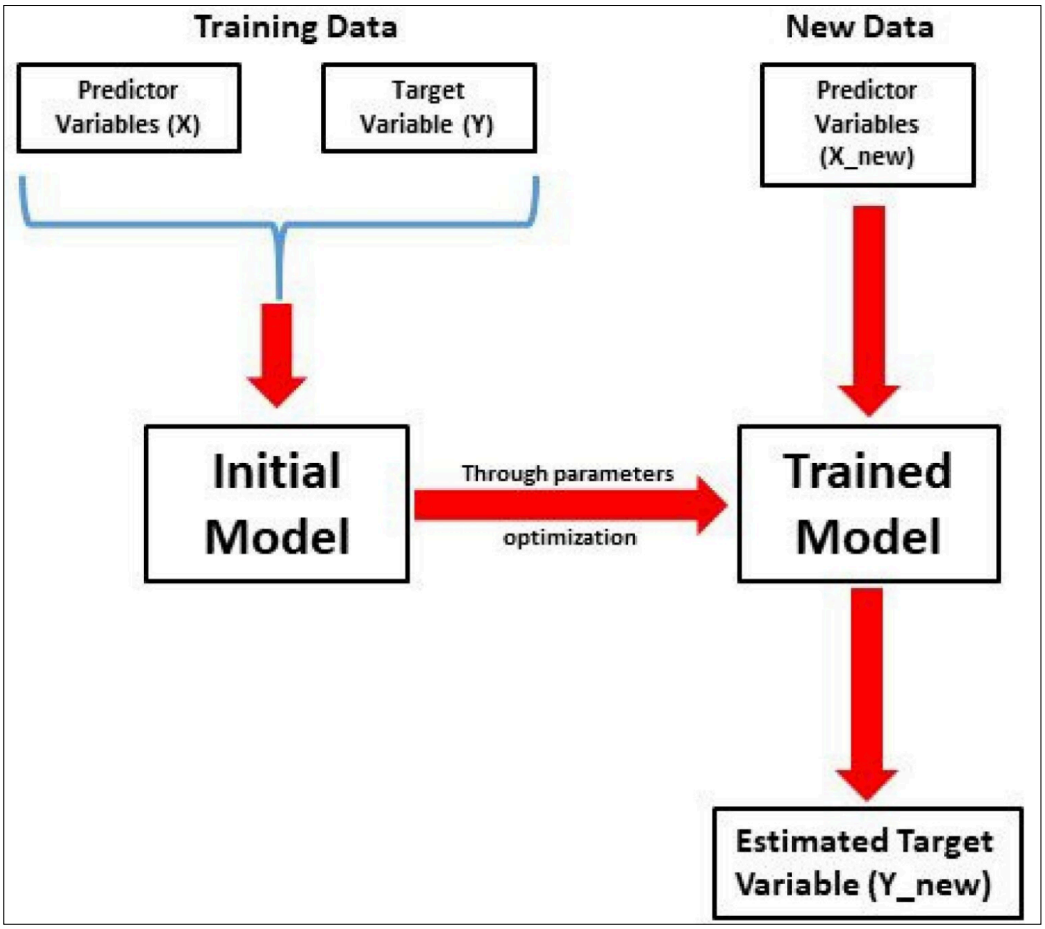


Figure 3.4: Conventional ML Workflow[143]

HML enhances the ML process by combining algorithms, methodologies, and procedures from different knowledge or application areas. This works out nicely. No one size will fit all. Because of this, ML is unable to solve all aspects of a problem. Some methods can handle noisy data but may require assistance navigating complex feed space. HML can complement the candidate approaches and overcome their drawbacks under specific settings. Figure 3.5 shows the HML workflow framework.

The options for hybridising classic ML methods are virtually limitless, which may be done for every one of them to construct new hybrid models in various ways. This essay will discuss architectural integration, data manipulation, and optimising model parameters.

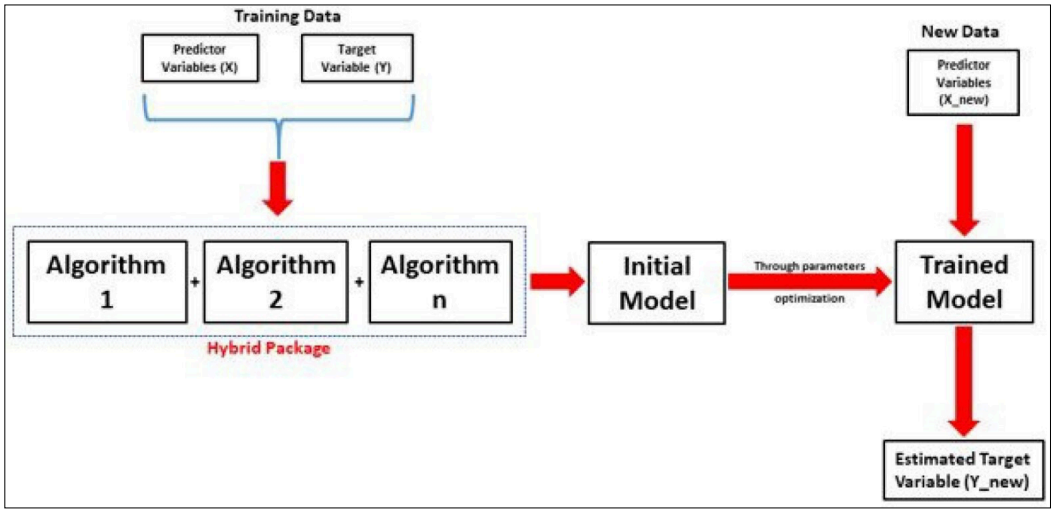


Figure 3.5: HML Workflow[143]

The goal of this category of HML is to build a freestanding algorithm that is more robust by combining the architecture of two or more conventional algorithms, totally or partially, in a complementary manner. This type of HML smoothly combines the algorithms. The adaptive neuron-fuzzy inference system, or ANFIS, is the model that is most frequently used as an example. ANFIS has been implemented throughout the years and is typically considered a traditional machine-learning approach that stands on its own. It is a hybrid approach that takes elements from both ANN and fuzzy logic. The architecture of ANFIS is made up of five different layers, as can be seen in Figure 3.6. The first three are from fuzzy logic, and the final two come from artificial neural networks (ANN).

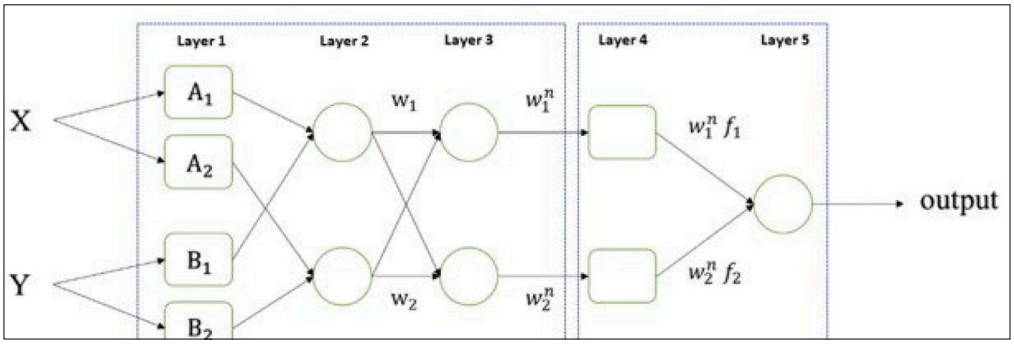


Figure 3.6: ANFIS Architecture [143]

3.9 Performance Metrics

In machine learning, it is essential to precisely evaluate the developed model to guarantee that the predictions accurately characterise the phenomenon intended to be described (for example, the prediction of disease, the estimation of future costs, etc.).

However, data scientists have access to various performance indicators (such as accuracy, precision, and recall), so it can take time to decide which to employ.

However, picking the appropriate metric for a particular model is essential to measure the model's performance objectively and in the proper context.

The term "Confusion Matrix" is a table that organises the four different values.

Accuracy

Accuracy is our model's proportion of true predictions. The Accuracy can range anywhere from 0 to 1. These two extremes equate to either entirely missing the predictions or always having the predictions spot on. For example, if our model can make accurate predictions, it will not generate false positives or negatives. This will result in the numerator and the denominator having the same value, increasing the Accuracy to 1.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad \text{..... (i)}$$

If, on the other hand, our system consistently makes inaccurate predictions, the number of True Positives and True Negatives will both be zero. This will cause the equation to read zero divided by something positive, resulting in zero Accuracy.

Accuracy's value can lie between 0.5 and 1. However, if it is below 0.5, we can quickly improve our Accuracy by switching the order in which the predictions are labelled.

Precision

Data scientists typically rely on precision, recall, and specificity to overcome accuracy limitations. The term "precision" refers to the degree to which optimistic predictions were, in fact, accurate. To accomplish this, it counts the samples

predicted correctly as positive (TP) and then divides that number by the overall number of optimistic predictions, whether accurate or incorrect (TP, FP).

$$Precision = \frac{TP}{(TP + FP)}$$	..... (ii)

Recall(Sensitivity)

Along the same lines as Precision, the purpose of the Recall metric is to determine what percentage of actual positives were recognised accurately. The calculation involves dividing the number of correctly predicted positive samples by the total number of positive samples, including true positives (TP) and false negatives (FN).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$	..... (iii)

Specificity

Specificity is calculated in a manner that is symmetrical to that of Recall, also called Sensitivity. This is done by dividing the amount of accurately predicted negative samples by the total number of negative phrases that were either precisely predicted as negative phrases or incorrectly predicted as positives (TN, FP). Specificity aims to determine what percentage of actual negatives was identified correctly.

$$Specificity = \frac{TN}{(FP + TN)}$$	..... (iv)

AUC

One of the problems with Accuracy is that it can result in an artificially inflated performance if the distribution of the classes is not particularly well balanced. This is one of the concerns. The” AUC"- Area Under Curve (we will talk more about this metric in the following paragraphs), is a measurement that determines the whole two-dimensional area that lies underneath the entire ROC curve.

It is a performance measure aggregated over many feasible classification thresholds. This is interpreted as the chance that the model places a random positive sample higher on its list than a random negative sample. Even though it can only be used in binary classification scenarios (i.e., not with more than two classes as target), AUC is a significant parameter, especially when dealing with imbalanced classes, and it is one of the most often used performance metrics in classification.

F1 score

The F1 score is a performance metric that calculates the harmonic mean of the recall and precision values. This is a numerical measure that goes from 0 to 1. A score of 1 shows perfect Precision and Recall, whereas a score of 0 means that neither Precision nor Recall can be measured. The lowest possible value is 0, which can only be achieved if neither Precision nor Recall can be measured.

$$F1\ Score = \frac{2TP}{(2TP + FP + FNN)} \quad \text{..... (v)}$$

Due to the utilisation of numerous layers of neural networks, "Deep Learning" systems necessitate less human intervention than "traditional" learning systems. These networks progressively acquire knowledge from their mistakes and shortcomings, thereby diminishing the need for human supervision. Due to the input database's significant impact on the model's accuracy, deep learning was chosen for this experiment instead of machine learning.

When dealing with a massive data collection, it is strongly suggested that you utilise DL designs. Furthermore, it is determined by the quality of the information delivered when being trained. When the process of feature engineering is carried out efficiently, machine learning models have the potential to produce merely satisfactory outcomes, even when working with minimal data sets.

Researchers, e-commerce platforms, and internet businesses have a formidable obstacle in detecting counterfeit reviews. Based on our research, the existing text-generating systems generate deceptive evaluations that closely resemble genuine assessments, making it challenging for individuals to distinguish between the two.

This study comprehensively examines the current state of detecting fraudulent reviews and employing machine and deep learning methodologies. Initially, we have examined the multitude of feature extraction methodologies researchers have used. Subsequently, we comprehensively analysed the existing datasets, meticulously documenting the methods employed in their creation. Subsequently, we comprehensively examined multiple models for detecting fraudulent reviews, employing summary tables. However, deep learning can enhance performance by utilising the hybrid modelling paradigm. Our primary goal for future research is to focus on developing text-enhancing columns. We will improve the existing dataset in Phase II by incorporating a polarity variable. Subsequently, we will employ ensemble modelling for novelty.

Based on their performance in the literature study, we picked ten machine learning models and five deep learning models for prescriptive analytics. Initially, we looked at the base models of classification problems using naive Bayes, logistic regression, and k-nearest neighbours. Then, we choose SGD, which stands for Stochastic Gradient Descent. "Stochastic" means not planned. In each step, SGD doesn't use the whole collection to determine the gradient. Instead, it estimates the gradient from a small group of data points (called a "mini-batch").

This makes it fast to use with big datasets. XGBoost and CatBoost have also been added. Both are known for performing best on various classification jobs, such as finding fake reviews. They use ensemble methods, which combine predictions from several decision trees to make more reliable and accurate predictions than forecasts from a single model. Boosting models can choose features and do engineering automatically, and they can use regularisation methods to keep them from fitting too well.

We chose the best deep learning models for sequential data by processing data in both directions—that is, by sending it forward and backward—so that they could target the context of the data. Roberta and Afinn were the best models for this. For reviews, techniques should be thought about based on a literature study.

So finally, ten machine learning models (Gaussian, Multinomial, Bernoulli, Complement Naive Bayes Classifiers, k-nearest neighbours, Logistic Regression, Stochastic Gradient Descent (SGD), Random Forest, XGBoost, and CatBoost) were

used as a baseline for checking performance. Next, we used hybrid and Ensemble modelling in combination with deep learning models known as LSTM, Bidirectional LSTM, multi-dense LSTM, GRU, and Bidirectional GRU to detect fake reviews. OpenAI's concept of optimising the RoBERTa model for a particular task was a foundation for our FRARBiLSTM model.

Models	Why Selected or Not Selected based on features?
Gaussian Naive Bayes	Handles categorical features well,Simple and efficient**<Selected>**
Multinomial Naive Bayes	Efficient for large datasets with discrete features**<Selected>**
Bernoulli Naive Bayes	Can be good for identifying presence/absence of specific keywords**<Selected>**
Complement Naive Bayes	may not be ideal for standard fake/real review classification but can be used for identifying outliers**<Selected>**
K-nearest neighbours	Can capture local patterns in the data**<Selected>**
Logistic Regression	Provides insights into feature importance**<Selected>**
Stochastic Gradient Descent	Fast Training**<Selected>**
Random Forest	Provides good overall accuracy**<Selected>**
XGBoost	High accuracy and handles complex data**<Selected><Selected>**
CatBoost	Potentially faster training for large datasets with categorical features**<Selected>**
ANN	Not handle Sequential Data**<Not Selected>**
RNN	RNN works on short term memory/ sequence of reviews only.**<Not Selected>**
MLP	Very complex for Reviews Data**<Not Selected>**
CNN	For Image only**<Not Selected>**
LSTM, Multi-dense LSTM, Bi-LSTM	works on Long term memory, Short term memory, 3 gates(Forget,Input,Output),both forward and backward processing**<Selected>**
GRU, Bi-GRU	both forward and backward processing, Update Gate, faster**<Selected>**

Table 3.1: Models Selection based on Features

Chapter 4

Simulations, Results and Analysis

We have implemented the entire process of simulations, results, and analysis objective-wise, as mentioned in Chapter 3 under the proposed framework.

Objective 1 To study and evaluate the performance of existing models for Analytics on Online Social Networks and E-Commerce sites.

4.1 Experimental Setup

For simulation, we use Google Colab, numpy, sci-kit-learn, matplotlib, panda,TensorFlow, Python, GPU-based architecture and some additional neural network-based libraries .

4.2 Database Used

The primary objective should be compiling an extensive database comprising authentic and deceptive reviews. We chose a significant dataset due to its equitable nature. It should include an equitable distribution of genuine and fraudulent evaluations and be employed by excellent, most recent research papers. Several online platforms, such as OSF, Amazon, Yelp, and TripAdvisor, offer the dataset for download. The OSF Fake Review Dataset (https://osf.io/tyue9/) was utilised for this study. Figure 4.1 presents information on the fake reviews dataset. This dataset is the classification dataset in which we have three independent variables named category of the review, rating of product provided by the customer, text_ means reviews of customers, and one dependent variable named label with binary values. CG means computer-generated (fake-0) reviews, and OR means Original(genuine-1) reviews. So, the dataset contains 40432 rows and 4 columns.

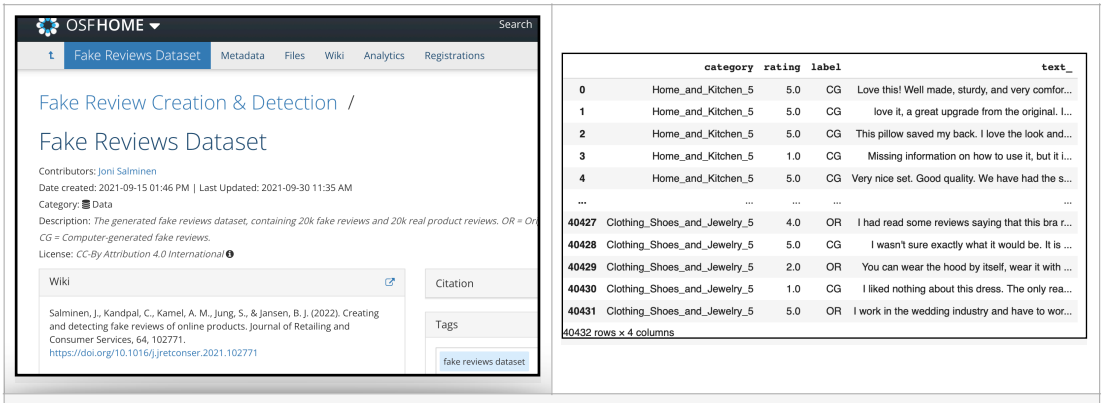


Figure 4.1 : Fake Review Dataset

4.3 Comparative study of Models

We execute experiments employing ten ML and five DL classifiers to select the optimal model. These classifiers are evaluated using a classification matrix and performance parameters, as shown in the tabular form below.

Classifiers (ML/DL)	Fake Review parameters	Feature Engineering	Performance parameters	Result	
				Orignal Dataset	Enriched Dataset (Objective 2)
GNB	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score	70.48	70.38
MNB	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score	78.95	77.19
BNB	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score	73.43	72.25
CNB	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score	78.96	77.19
KNN	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score	61.7	61.33
LR	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score	82.51	83.35
RF	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score	75.61	75.66
XGB	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF, Boosting	Accuracy , Precision,Recall , F1 score	82.51	82.3
CatB	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF, Boosting	Accuracy , Precision,Recall , F1 score	87.22	87
SGD	POS, LIWC, Bigram, BOW, Similarity Index	TF-IDF	Accuracy , Precision,Recall , F1 score, Loss Function	82.37	82.52
Bi-LSTM	POS, LIWC, Bigram, Contextual Features	AFINN	Training Accuracy, Precision,Recall , F1 score, Loss Function	99.9	99.9
LSTM	POS, LIWC, Bigram, Contextual Features	AFINN	**Training Accuracy,** Precision,Recall , F1 score, Loss Function	90	90.71

Classifiers (ML/DL)	Fake Review parameters	Feature Engineering	Performance parameters	Result	
				Orignal Dataset	Enriched Dataset (Objective 2)
GRU	POS, LIWC, Bigram, Contextual Features	AFINN	**Training Accuracy,** Precision,Recall , F1 score, Loss Function	96.12	96.24
Bi-GRU	POS, LIWC, Bigram, Contextual Features	AFINN	**Training Accuracy,** Precision,Recall , F1 score, Loss Function	98.79	98.78
Multi Dense LSTM	POS, LIWC, Bigram, Contextual Features	AFINN	**Training Accuracy,** Precision,Recall , F1 score, Loss Function	92	91.69

Table 4.1: Performance Evaluation of Existing Models

Accuracy 0.7038456782490417

	precision	recall	f1-score	support
CG	0.66	0.83	0.74	4047
OR	0.77	0.58	0.66	4040
accuracy			0.70	8087
macro avg	0.72	0.70	0.70	8087
weighted avg	0.72	0.70	0.70	8087

Figure 4.2(a) : GNB

	precision	recall	f1-score	support
CG	0.73	0.87	0.79	4047
OR	0.83	0.68	0.75	4040
accuracy			0.77	8087
macro avg	0.78	0.77	0.77	8087
weighted avg	0.78	0.77	0.77	8087

Figure 4.2(b) : MNB

Accuracy 0.7225176208730061

	precision	recall	f1-score	support
CG	0.67	0.88	0.76	4047
OR	0.83	0.56	0.67	4040
accuracy			0.72	8087
macro avg	0.75	0.72	0.72	8087
weighted avg	0.75	0.72	0.72	8087

Figure 4.2(c): BNB

Accuracy 0.7719797205391369

	precision	recall	f1-score	support
CG	0.73	0.87	0.79	4047
OR	0.83	0.68	0.75	4040
accuracy			0.77	8087
macro avg	0.78	0.77	0.77	8087
weighted avg	0.78	0.77	0.77	8087

Figure 4.2(d) :CNB

Accuracy 0.6133300358600222

	precision	recall	f1-score	support
CG	0.58	0.85	0.69	4047
OR	0.72	0.37	0.49	4040
accuracy			0.61	8087
macro avg	0.65	0.61	0.59	8087
weighted avg	0.65	0.61	0.59	8087

Figure 4.2(e) :KNN

Accuracy 0.8335600346234697

	precision	recall	f1-score	support
CG	0.82	0.85	0.84	4047
OR	0.84	0.82	0.83	4040
accuracy			0.83	8087
macro avg	0.83	0.83	0.83	8087
weighted avg	0.83	0.83	0.83	8087

Figure 4.2(f) : LR

Accuracy 0.7566464696426364

	precision	recall	f1-score	support
CG	0.75	0.76	0.76	4047
OR	0.76	0.75	0.76	4040
accuracy			0.76	8087
macro avg	0.76	0.76	0.76	8087
weighted avg	0.76	0.76	0.76	8087

Figure 4.2(g) : RF

Accuracy 0.8252751329293928

	precision	recall	f1-score	support
0	0.79	0.88	0.84	4047
1	0.87	0.77	0.81	4040
accuracy			0.83	8087
macro avg	0.83	0.83	0.82	8087
weighted avg	0.83	0.83	0.82	8087

Figure 4.2(h) : SGD

Figure 4.2: Models Analysis -Precision Vs. Recall Vs. f1-score

	precision	recall	f1-score	support
0	0.92	0.90	0.91	4016
1	0.90	0.92	0.91	4071
accuracy			0.91	8087
macro avg	0.91	0.91	0.91	8087
weighted avg	0.91	0.91	0.91	8087

Figure 4.3(a):XGBoost with RoBERTa

	precision	recall	f1-score	support
0	0.92	0.91	0.91	4016
1	0.91	0.92	0.92	4071
accuracy			0.92	8087
macro avg	0.92	0.92	0.92	8087
weighted avg	0.92	0.92	0.92	8087

Figure 4.3(b): CatBoost with RoBERTa

Figure 4.3: Ensemble Models: XGBoost Vs. CatBoost

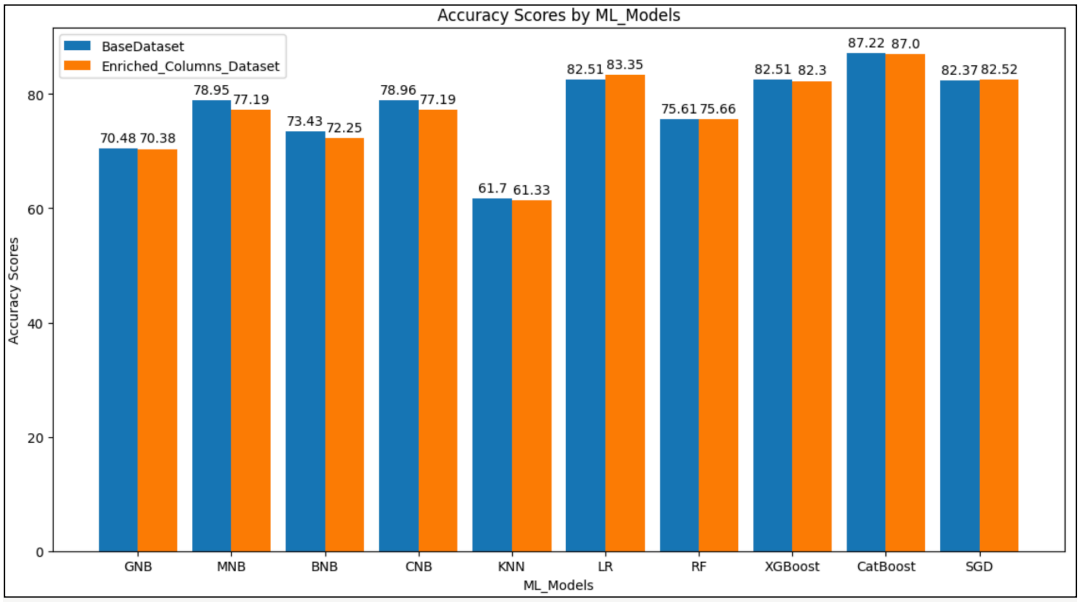


Figure 4.4 : ML Models Results over Base dataset Vs. Enriched Dataset

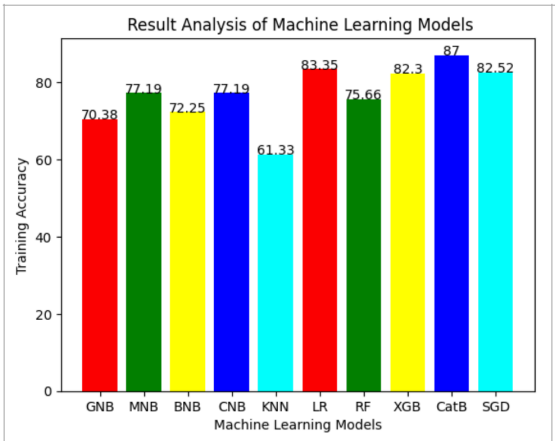


Figure 4.5(a): Results Analysis of Ml Models

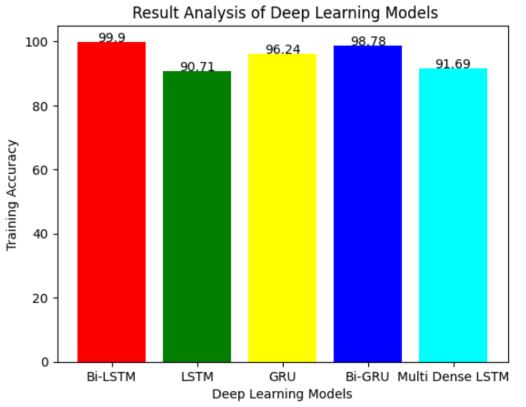


Figure 4.5(b): Results Analysis of DL Models

Figure 4.5 : Comparative Analysis of ML and Deep Learning Models

We conclude that either we used the original or enriched dataset, the BiLSTM model trained best among all the traditional and deep learning models. So, we have to choose **BiLSTM as a base for a proposed model** that will integrate with additional features.

Objective 2:To extract different patterns of unstructured data from Online Social Network pages and E-commerce sites through Analytics tools by examining customer reviews

4.4 AFINN and RoBERTa

4.4.1Enriching text columns: AFINN

We suggest utilizing the AFINN lexicon to augment the sentiment polarisation of textual sections. The AFINN lexicon is a compilation of English idioms that were assessed for their emotional valence by Finn Rup Nielsen between 2009 and 2011. Each term was given an integer value to indicate its degree of positivity or negativity. However, those multi-word phrases are not included here because they are found in the original lexicon. The Open Database Licence version 1.0 (ODbL) shares the original lexicon with the public. If you customise the dictionary in any way, you must keep the customised lexicon open-source and use a license analogous to the original.

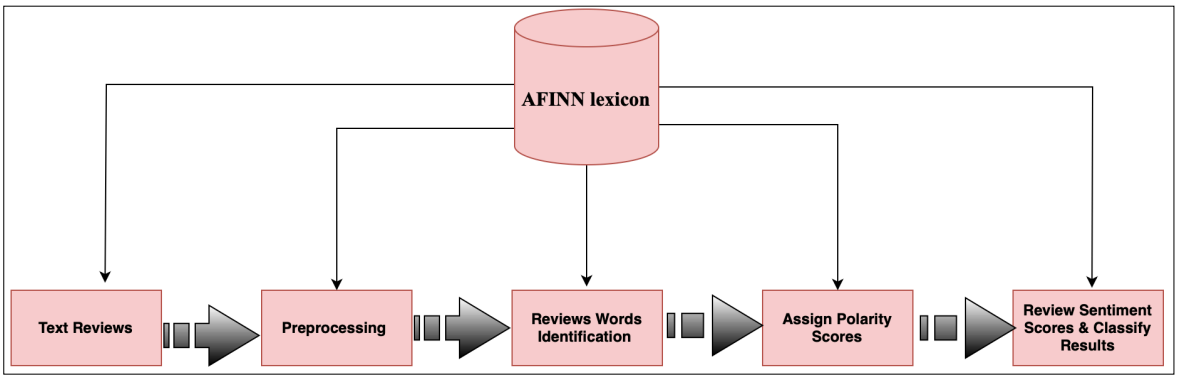


Figure 4.6: Functionality of AFINN lexicon

Because of its potential financial benefits, sentiment analysis (also known as SA) has recently emerged as one of the most fascinating issues in text analysis. When reading opinion reviews, one of the most significant obstacles that SA needs to overcome is to distinguish feelings from people's perspectives and distinguish between evaluations that are either positive or intentionally negative. In addition, the user reviews gathered can be separated into two categories: good and evil. Consumers may utilise either type of review to help them choose a product.

Using machine learning techniques, this study's authors intend to categorise reviews into categories according to whether they are excellent or negative. The information that may be found after the Afinn lexicon as an enriched data column is shown in Figure 4.7.

	category	rating	label	text_	sent_score	sent_category	ENRICHED_TEXT
0	Home_and_Kitchen_5	5.0	0	Love this! Well made, sturdy, and very comfor...	9.0	positive	This review is positive and has a rating of fi...
1	Home_and_Kitchen_5	5.0	0	love it, a great upgrade from the original. I...	6.0	positive	This review is positive and has a rating of fi...
2	Home_and_Kitchen_5	5.0	0	This pillow saved my back. I love the look and...	5.0	positive	This review is positive and has a rating of fi...
3	Home_and_Kitchen_5	1.0	0	Missing information on how to use it, but it i...	1.0	positive	This review is positive and has a rating of on...
4	Home_and_Kitchen_5	5.0	0	Very nice set. Good quality. We have had the s...	8.0	positive	This review is positive and has a rating of fi...
...	...	...	...	...	...	...	...
40427	Clothing_Shoes_and_Jewelry_5	4.0	1	I had read some reviews saying that this bra r...	22.0	positive	This review is positive and has a rating of fo...
40428	Clothing_Shoes_and_Jewelry_5	5.0	0	I wasn't sure exactly what it would be. It is ...	61.0	positive	This review is positive and has a rating of fi...
40429	Clothing_Shoes_and_Jewelry_5	2.0	1	You can wear the hood by itself, wear it with ...	2.0	positive	This review is positive and has a rating of tw...
40430	Clothing_Shoes_and_Jewelry_5	1.0	0	I liked nothing about this dress. The only rea...	62.0	positive	This review is positive and has a rating of on...
40431	Clothing_Shoes_and_Jewelry_5	5.0	1	I work in the wedding industry and have to wor...	17.0	positive	This review is positive and has a rating of fi...

40432 rows × 7 columns

Figure 4.7: Fake Reviews Enriched_Text Dataset

During the preprocessing stage, the datasets were preprocessed to remove the noise, which included things like stop words, URLs, emojis, and other similar items. The NLTK toolkit1, an open-source package frequently employed, was utilized to perform the preprocessing. After tokenization, we proceeded to eliminate the stop words that were causing interference with the text categorization process.

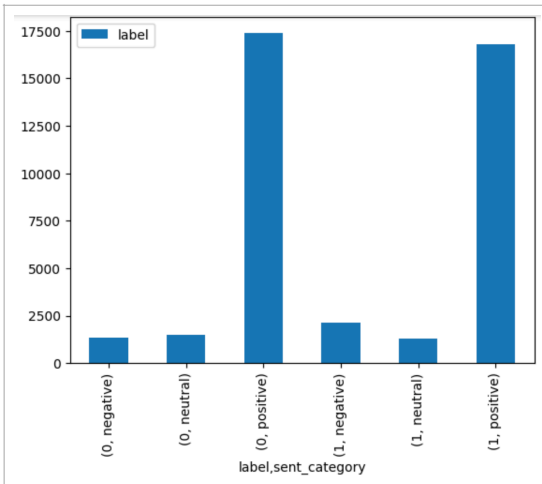


Figure 4.8(a): Sentiment Extraction label and category-wise

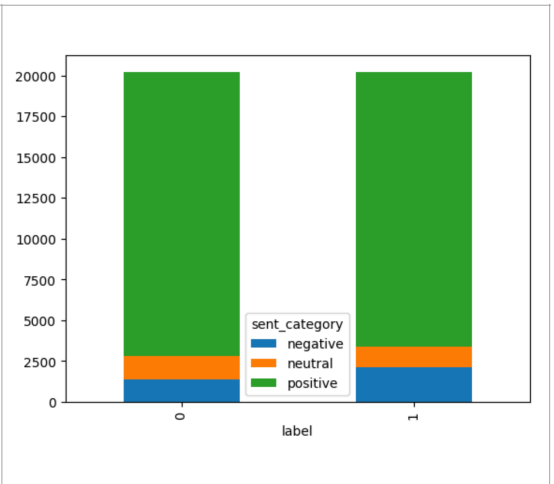


Figure 4.8(b): Sentiment Extraction label-wise

Figure 4.8: Sentiment Extraction

A sample size of 40,432 balanced reviews (20,216 fake and 20,216 original) and an 80/20 training-testing split should reduce overfitting in our fake review detection research. Over 40,000 data points are adequate to show actual and false review patterns. This lets the model learn complex feature-target variable correlations (fake/ honest review). Classification jobs require equal numbers of false and authentic evaluations. This prevents imbalanced datasets from biassing the model toward the dominant class. Machine learning often uses an 80/20 train-test split. It gives 80% of the data needed to train the model and 20% for unbiased judgement on unseen data.

This divide prevents **overfitting** by learning the model with training data and verifying its generalizability with testing data. The model must learn patterns that apply to new reviews rather than memorising training data. The setup is ideal; however, overfitting must be monitored during exercise. This can be avoided with **early stopping.**

4.4.2 RoBERTa

The dataset is loaded into a RoBERT to build word embeddings. Word embeddings are huge vector representations of the text words contained in the dataset. After that, the input is entered into the classification models so that they can begin their training. The results are assessed using a confusion matrix representation to determine performance metrics. The classifier that produces the best results is then saved and used later to determine whether user reviews are genuine. Roberta analyses input sequences and creates contextualized word representations.

One of the most important distinctions that can be made comparing Roberta and BERT is the fact that the earlier model had been trained on a significantly more extensive dataset and made use of a training strategy that was more effective. To train Roberta, we employed a text dataset that was 160 GB in size. This is more than ten times the dataset utilized during BERT's training. During the presentation of the model, Roberta makes use of dynamic masking. This assists the model in developing word representations that are more resilient and general. It has served as the basis for constructing several profitable NLP models and is utilized extensively in academic and commercial settings. A dependable and helpful language model is RoBERTa. It has advanced natural language processing (NLP) and other applications.

RoBERTa's designers made several fundamental design changes to its architecture and training method to improve the BERT architecture's results. BERT's architecture is nearly identical to RoBERTa's. These adjustments:

- During the training phase for the Next Sentence Prediction (NSP) objective, an algorithm gets the instruction to determine, with the use of an auxiliary NSP loss, regardless of if the detected content segments originated from the same document or other documents. This is done to determine whether or not the NSP objective would be successful. This step is taken to achieve what is commonly referred to as

the NSP goal. The writers carried out several tests, during which they either maintained or marginally improved the performance of downstream jobs by adding or removing NSP loss from various versions. The authors concluded that deleting NSP loss either maintains or marginally improves the performance of downstream tasks.BERT is initially configured to train for a batch size of 256 sequences and 1 million steps. This is its default setup.

- To conduct this inquiry, the investigators tested the model employing 125 stages of 2K patterns and 31K steps of 8K patterns of a single batch. This results in two positive effects: first, the difficulty of the masking language modelling objective is made more difficult as a direct result of the massive batches, and second, the accuracy of the end task is improved as a direct result of the increased difficulty of the original study. Using distributed parallel training makes parallelising huge batches possible and more straightforward. Making real-time adjustments to the masking pattern is feasible using the BERT architecture. This architecture produces A single static mask because the masking method is executed once during the stage devoted to the data preprocessing. To prevent the use of an identical stable mask, the data used for training is replicated and masked ten times, spanning a total of forty epochs. During every one of these ten iterations, a different masking method is implemented. This is done to prevent the usage of a single static mask. Consequently, the mask is utilized in only four distinct epochs.

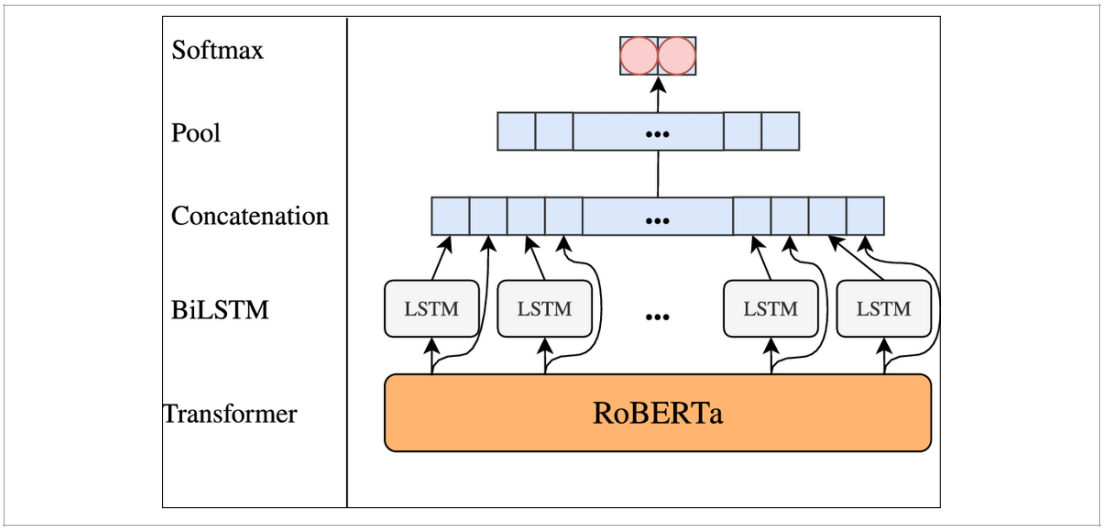


Figure 4.9: RoBERTa Architecture [142]

- This method is distinguished from dynamic masking, in which new masking is created for each new data input into the model by its capacity to supply a distinct masking set at any given time. In dynamic masking, new masking is created for each data set added to the model.

Objective 3:To identify the various parameters for the recognition of fake reviews

4.5: POS, LIWC, Bigram integrated with RoBERTa

Fake review parameters are divided into reviewer, review and Network-centric. We have to focus on review-centric parameters in which ensemble POS, LIWC, and Bigram with RoBERTa dynamic embedding techniques so that the classification of a review is either spam or ham based on the bidirectional context of the sentences.

POS	Count	Percentage
Noun	1743	23.72%
Verb	1370	18.52%
Adjective	1194	16.18%
Adverb	817	11.12%
Pronoun	594	8.16%
Conjunction	444	6.08%
Article	294	4.04%
Preposition	221	3.06%

Figure 4.10(a):POS

LIWC	Count	Percentage
I	1341	18.94%
We	631	9.02%
You	558	8.13%
They	408	6.03%
This	393	5.82%
That	343	5.07%
It	339	4.99%
Was	329	4.85%
A	292	4.29%

Figure 4.10(b):LIWC

LIWC	Count	Percentage
Positive	2253	31.72%
Negative	1305	18.70%
Social	1094	15.78%
Affective	1042	15.18%
Cognitive	725	10.32%
Perceptual	386	5.56%
Bodily	301	4.44%
Time	229	3.22%
Other	140	1.98%

Figure 4.10(c):LIWC

Bigram	Count	Percentage
the product	1022	14.80%
this product	789	11.40%
i love	725	10.40%
very good	695	10.00%
amazing product	673	9.60%
great product	651	9.20%
love this product	637	9.00%
highly recommend	593	8.40%
would recommend	559	7.90%

Figure 4.10(d): Bigram

Figure 4.10: POS, LIWC, Bigram on OSF dataset

4.5.1 Role of POS(Part of Speech)

Nouns, verbs, and adjectives are the most common POS tags in OSF false reviews. OSF fake reviews generally utilise positive language to promote a product or service.

OSF fake reviews also use adverbs and pronouns less often than nouns, verbs, and adjectives. OSF fake reviews use the fewest conjunctions, articles, and prepositions.

POS tags can reveal a review's content and whether it's fraudulent. *A review with more positive nouns, verbs, and adjectives is likelier to be dishonest than one with a more balanced mix of POS tags.*

4.5.2 Role of LIWC(Language Inquire Word Count)

"I" is the most common LIWC term in OSF false reviews, followed by "we," "you," and "they." This is because OSF false reviews are generally written in the first person as if the reviewer is utilizing the product or service.

"I," "we," "you," and "they" are more common in OSF false reviews than "this," "that," "it," and "was." "A" and "the" are the least common LIWC words in OSF fake reviews.

Analyzing a review's LIWC terms can help determine its authenticity and content. For instance, a review with many first-person pronouns is more likely to be fraudulent than one with a more balanced LIWC word mix.

Positive, negative, and social terms are the most prominent LIWC categories in OSF false reviews. OSF fake reviews generally utilize positive language to promote a product or service.

Positive, negative, and social words dominate affective, cognitive, and perceptual words in OSF false reviews. OSF fake reviews rarely use bodily, time, or other LIWC categories.

A review's LIWC categories can help determine its authenticity and content. A review with many favourable and social terms is more likely to be fraudulent than one with a balanced LIWC mix.

Of fact, LIWC sometimes catches false reviews. It can be used with other methods like POS labelling and BOW, BERT, and RoBERTa analysis.

4.5.3 Role of BiGram

As you can see, OSF false evaluations most often use "the product," followed by "this product," "I love," and "very good." This is because OSF fake reviews generally emphasize the product or service to attract readers.

The bigrams "amazing product," "great product," and "love this product" are also popular in OSF false reviews, but they are less common than "the product," "this product," and "I love". "Would recommend" and "Highly recommend" are the least common bigrams in OSF fake reviews.

Analysing the bigrams of a review can help determine its authenticity and content. For instance, a review with many bigrams emphasising the product or service is likelier to be fake.

Of course, bigrams can't always spot fake reviews. They can be used with other methods like POS tagging and LIWC analysis.

Objective 4 : To propose a Fake Review Authentication model based on extracted customer-review patterns.

4.6: FRARBiLSTM:(Without POS, LIWC, Bigram)

As can be seen in Figure 3.1, our given model has been broken down into its constituent parts. The information has been cleaned using natural language processing (NLP) techniques like stemming and removing punctuation and stop words. Stemming and reducing them to lowercase English letters are two further natural language processing techniques. Figure 4.11 describes the processes performed to standardise the data. In the second stage, data is embedded, and text is enhanced with sentiment polarity to ensure that the feature selection process uses a high-quality dataset. Using Word Embedding techniques, such as Roberta, this method will mathematically represent the text.

Figure 4.11 illustrates the dataset's properties containing fake reviews by demonstrating that both classes are proportionately represented. The classification is accurate because each category has an equal number of samples. There is no need to take into account any mismatch when choosing a technique or modifying the data. The quantity of evaluations across the most popular product categories is generally similar, as depicted in Figure 4.12.

As a result, the classification algorithm has a high probability of acquiring the ability to distinguish between different classes in diverse product categories, hence enhancing its capacity to generalize.

The EDA process is an approach that includes aggregating the data with the help of statistical techniques and visualisation tools to bring the most significant elements of the data into focus for further investigation. To accomplish this, the dataset must be analysed from various perspectives and characterised, and its findings must be described without any assumptions about the information the dataset contains.

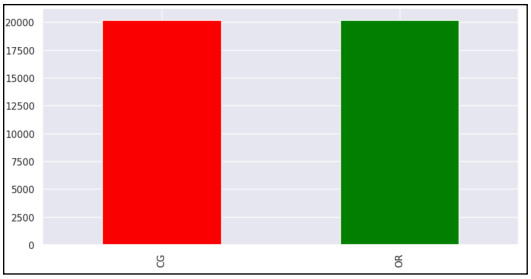


Figure 4.11(a): Frequency Distribution of Output Variable

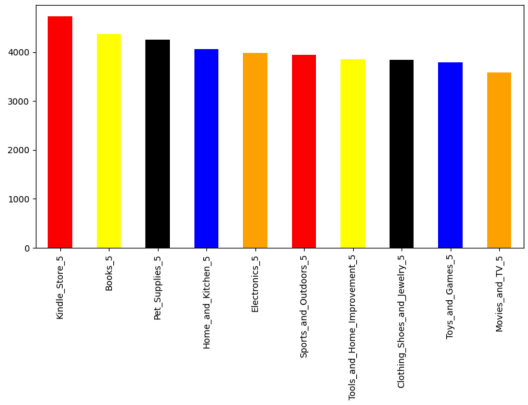


Figure 4.11(b): Reviews Count by Category

Figure 4.11: Exploratory Data Analysis

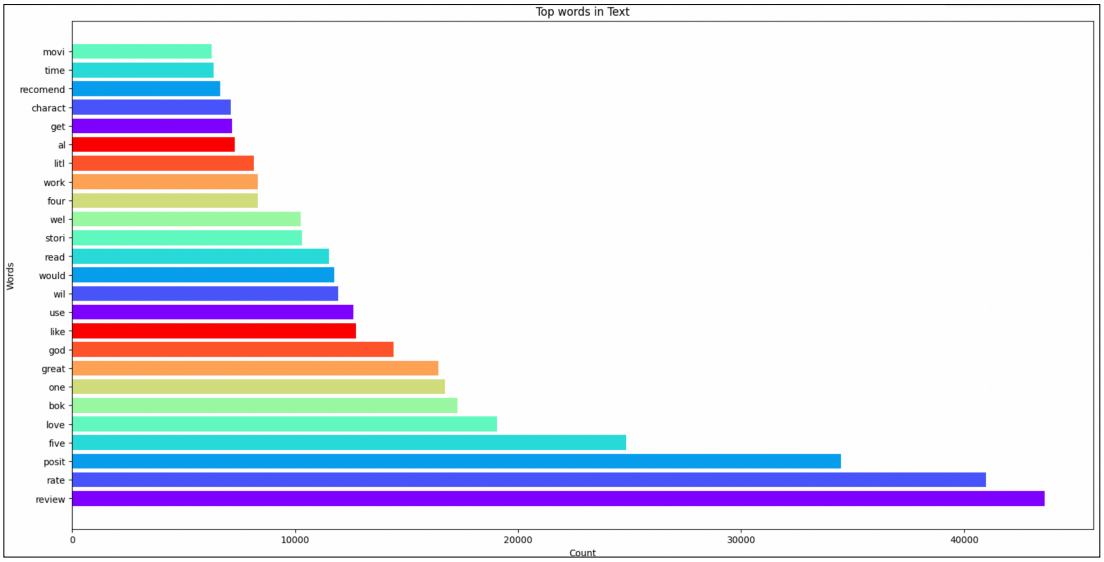


Figure 4.12: Top Words in Reviews

A word cloud is a graphical depiction illustrating the frequency and significance of particular terms in a text. This is demonstrated in Figure 4.13 and Figure 4.14. The arrangement of the words can frequently be arbitrary, with the magnitude of each phrase being proportional to the frequency of its occurrence in the text.

Words with lower frequency are typically displayed in smaller font sizes, while words with higher frequency are usually given in larger font sizes. Text analysis frequently uses WC because these tools offer a concise and understandable summary of the information in a document and draw attention to the subjects and keywords of the utmost significance.

Word clouds can be generated by utilising many tools of software applications and internet utilities. Once developed, word clouds can be personalised by applying multiple colours, geometric shapes, and fonts to make them more aesthetically pleasing.

Figure 4.13: Label 0-Word Cloud

Figure 4.14: Label 1-Word Cloud

Diagram 4.15 represents the algorithm process of FRARBiLSTM in which, firstly, we load data, implement preprocessing operations, data shuffling, and split data for training and testing. After that, we load our FRARBiLSTM model. Then, we implement 50 epochs to enhance accuracy. Our model also uses backward and

forward propagation, implementing accuracy loss, validation loss, RMSE, accuracy, precision, recall, and F1 Score.

	category	rating	label	text_	sent_score	sent_category	ENRICHED_TEXT	Token
0	Home_and_Kitchen_5	5.0	0	Love this! Well made, sturdy, and very comfor...	9.0	positive	This review is positive and has a rating of fi...	[review, posit, rate, five, love, wel, made, s...
1	Home_and_Kitchen_5	5.0	0	love it, a great upgrade from the original. I...	6.0	positive	This review is positive and has a rating of fi...	[review, posit, rate, five, love, great, upgra...
2	Home_and_Kitchen_5	5.0	0	This pillow saved my back. I love the look and...	5.0	positive	This review is positive and has a rating of fi...	[review, posit, rate, five, pilow, save, back,...
3	Home_and_Kitchen_5	1.0	0	Missing information on how to use it, but it i...	1.0	positive	This review is positive and has a rating of on...	[review, posit, rate, one, mise, inform, use, ...
4	Home_and_Kitchen_5	5.0	0	Very nice set. Good quality. We have had the s...	8.0	positive	This review is positive and has a rating of fi...	[review, posit, rate, five, nice, set, god, qu...

Figure 4.15 (b): Preprocess data

Algorithm 1: Deep Learning Process for Spam Review Detection

```
loadData();
preprocessData();
shuffleData();
splitData();
wordEmbedding();
loadModel();
foreach epoch in epochNumber do
    foreach batch in batchSize do
        logit = model(feature);
        loss = crossEntropy(logit, target);
        loss.backward();
        Evaluation(trainData, model, bestTrainAccuracy);
        Evaluation(testData, model, bestTestAccuracy);
    end
end
Function Evaluation(data, model, best):
    correct = 0;
    foreach batch in batchSize do
        logit = model(feature);
        loss = crossEntropy(logit, target);
        correct += (predict.data == target.data ).sum();
    end
    accuracy = (correct/totalData)*100;
    best = max(best, accuracy);
return
```

Figure 4.15 (a): Algorithm

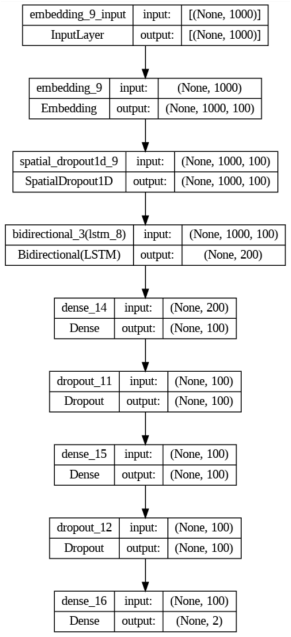


Figure 4.15 (c): Feature Embedding

Figure 4.15: Fake Review Detection Algorithm

The embedding layer, dense layer, output shape, dropout layer, and params used by our model FRARBiLSTM compared with the rest of the deep learning models without integrating POS, LIWC, Bigram, and Roberta as shown in Figure 4.16 to Figure 4.20, respectively. Figures 4.21–4.30 also show the verification, training accuracy, and loss graphs.

```
Model: "sequential_7"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_7 (Embedding)      (None, None, 64)          640000
_________________________________________________________________
lstm_6 (LSTM)                (None, None, 64)          33024
_________________________________________________________________
lstm_7 (LSTM)                (None, 16)                5184
_________________________________________________________________
dense_16 (Dense)             (None, 64)                1088
_________________________________________________________________
dropout_7 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_17 (Dense)             (None, 1)                 65
=================================================================
Total params: 679,361
Trainable params: 679,361
Non-trainable params: 0
```

Figure 4.16: LSTM

```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_4 (Embedding)      (None, None, 64)          640000
_________________________________________________________________
bidirectional_4 (Bidirection (None, None, 128)         66048
_________________________________________________________________
bidirectional_5 (Bidirection (None, 32)                18560
_________________________________________________________________
dense_10 (Dense)             (None, 64)                2112
_________________________________________________________________
dropout_4 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_11 (Dense)             (None, 1)                 65
=================================================================
Total params: 726,785
Trainable params: 726,785
Non-trainable params: 0
_________________________________________________________________
```

Figure 4.17:FRARBiLSTM

```
Model: "sequential_6"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_6 (Embedding)      (None, None, 64)          640000
_________________________________________________________________
gru_6 (GRU)                  (None, None, 64)          24960
_________________________________________________________________
gru_7 (GRU)                  (None, 16)                3936
_________________________________________________________________
dense_14 (Dense)             (None, 64)                1088
_________________________________________________________________
dropout_6 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_15 (Dense)             (None, 1)                 65
=================================================================
Total params: 670,049
Trainable params: 670,049
Non-trainable params: 0
```

Figure 4.18: GRU

```
Model: "sequential_5"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_5 (Embedding)      (None, None, 64)          640000
_________________________________________________________________
bidirectional_6 (Bidirection (None, None, 128)         49920
_________________________________________________________________
bidirectional_7 (Bidirection (None, 32)                14016
_________________________________________________________________
dense_12 (Dense)             (None, 64)                2112
_________________________________________________________________
dropout_5 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_13 (Dense)             (None, 1)                 65
=================================================================
Total params: 706,113
Trainable params: 706,113
Non-trainable params: 0
```

Figure 4.19: GRU Bidirectional

```
Model: "sequential_8"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_8 (Embedding)      (None, None, 64)          640000
_________________________________________________________________
lstm_8 (LSTM)                (None, None, 64)          33024
_________________________________________________________________
lstm_9 (LSTM)                (None, 16)                5184
_________________________________________________________________
dense_18 (Dense)             (None, 64)                1088
_________________________________________________________________
dense_19 (Dense)             (None, 128)               8320
_________________________________________________________________
dense_20 (Dense)             (None, 128)               16512
_________________________________________________________________
dropout_8 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_21 (Dense)             (None, 1)                 129
=================================================================
Total params: 704,257
Trainable params: 704,257
Non-trainable params: 0
```

Figure 4.20: Multi-Dense LSTM

According to the data, out of the these distinct DL models examined, including LSTM, FRARBiLSTM, Multi-dense LSTM, GRU, and Bidirectional GRU, the FRARBiLSTM model surpasses the rest of the classifiers in terms of accuracy the best (Epochs=50, Training and validation).

Models	Accuracy	Loss	RMSE Values
FRARBiLSTM	**99.9**	**0.001**	**0.031623**
LSTM	90.71	0.20	0.447214
GRU	96.24	0.37	0.608276
Bidirectional GRU	98.78	0.04	0.2
Multi Dense LSTM Model	91.69	0.45	0.67082

Table 4.2: Performance Results for Training Data

Models	Accuracy	Loss	RMSE Values
FRARBiLSTM	**85.59**	**1.82**	**1.349074**
LSTM	80.62	0.50	0.707107
GRU	88.00	0.61	0.781025
Bidirectional GRU	78.90	1.30	1.140175
Multi Dense LSTM Model	84.62	0.59	0.768115

Table 4.3: Performance Results for Testing Data

Models	Categories	Precision	Recall	F1 Score
FRARBiLSTM	**0**	**77**	**72**	70
	1	**89**	**86**	87
LSTM	0	66	71	78
	1	88	85	87
GRU	0	85	70	63
	1	73	90	74
Bidirectional GRU	0	79	75	78
	1	84	86	85
Multi Dense LSTM Model	0	96	76	46
	1	96	76	85

Table 4.4: CategoryWise Performance Results

4.6.1: Graphs

a) FRARBiLSTM- Figure 4.21 shows that the correlation between training accuracy and validation accuracy is minimal. This means the model may apply what it learned from the training data to novel data. How well a model is learning data patterns is depicted in Figure 4.22 as a function of training loss. Here, as validation loss goes down over time, we have a promising indicator.

Figure 4.21: FRARBiLSTM Accuracy Graph

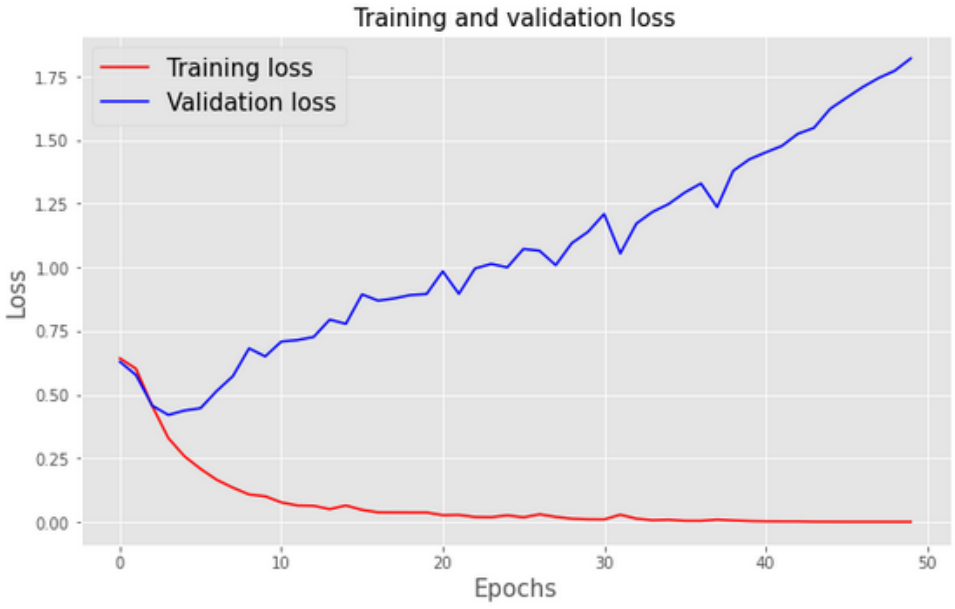


Figure 4.22: FRARBiLSTM Loss Graph

b) LSTM- Figure 4.23 demonstrates that LSTM achieved higher training accuracy of up to 90% and validation accuracy of up to 80%, both lower than the FRARBiLSTM model. Our training and validation loss is higher in Fig. 4.24 compared to the suggested model.

Figure 4.23: LSTM Accuracy Graph

Figure 4.24: LSTM Loss Graph

c) GRU- Figure 4.25 shows that GRU came exceptionally near the FRARBiLSTM model, with a training accuracy of up to 96% and a validation accuracy of up to 88%. Since there is a large gap between the training loss and the validation loss, as seen in Figure 4.26, further epochs will likely be required.

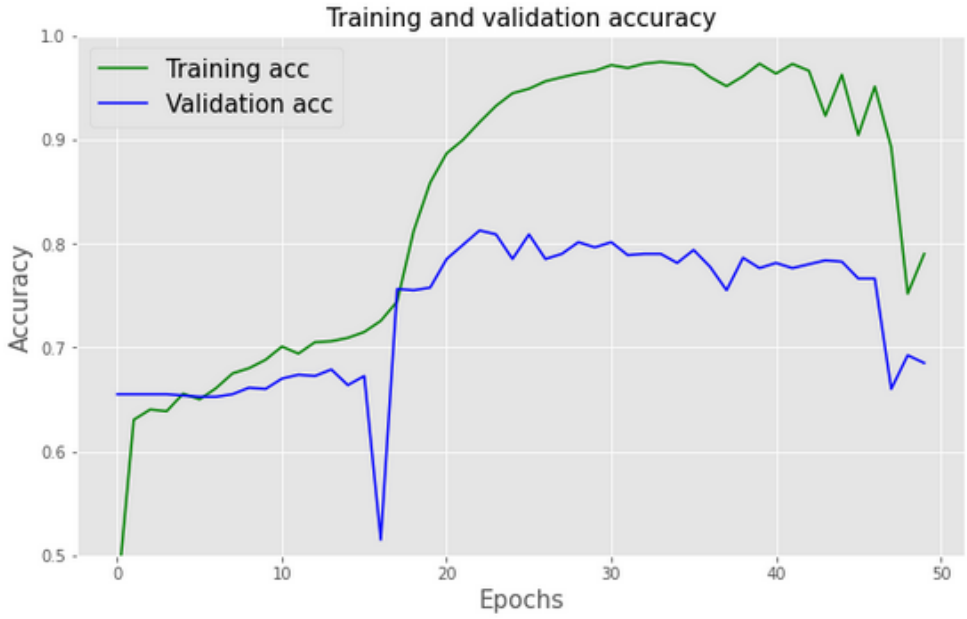


Figure 4.25: GRU Accuracy Graph

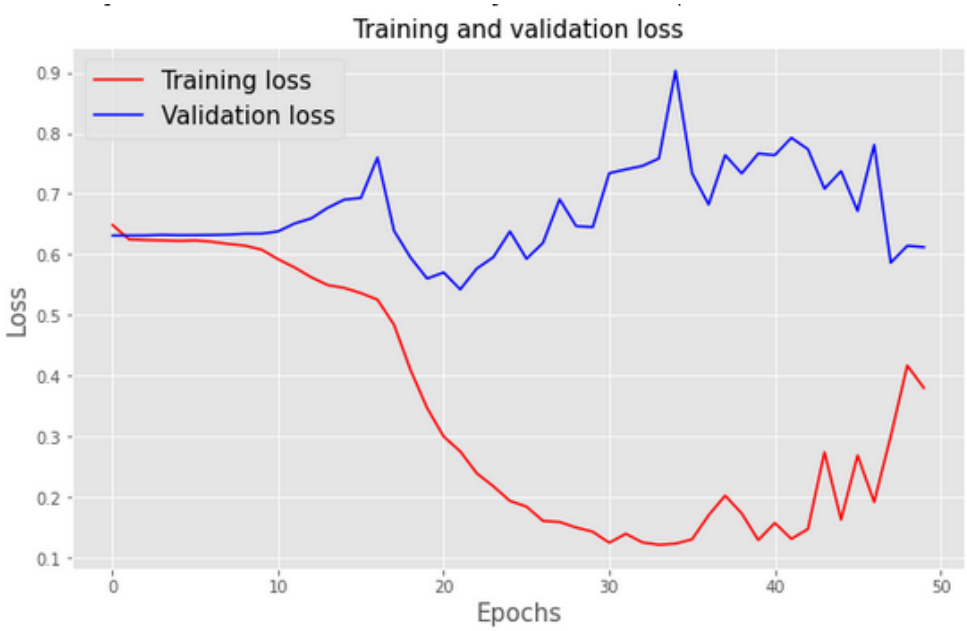


Figure 4.26: GRU Loss Graph

d) GRU Bidirectional- As shown in Figure 4.27, Bi-GRU's training accuracy reached 98%, while its validation accuracy reached up to 78%, which could be better compared to the FRARBiLSTM model. In addition, as seen in Figure 4.28, there is a sizeable discrepancy between the training loss and the validation loss, suggesting that additional epochs will be necessary.

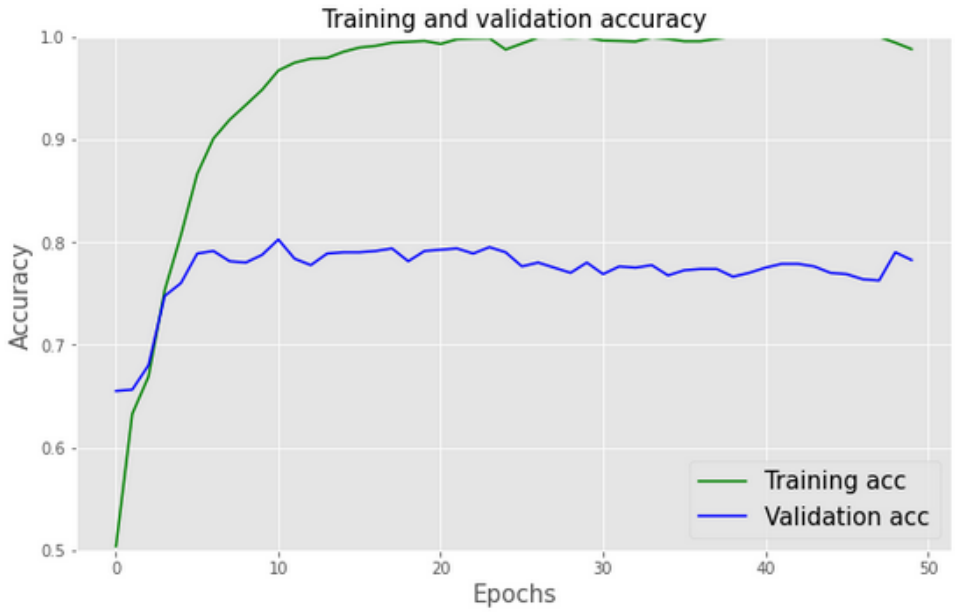


Figure 4.27: GRU Bidirectional Accuracy Graph

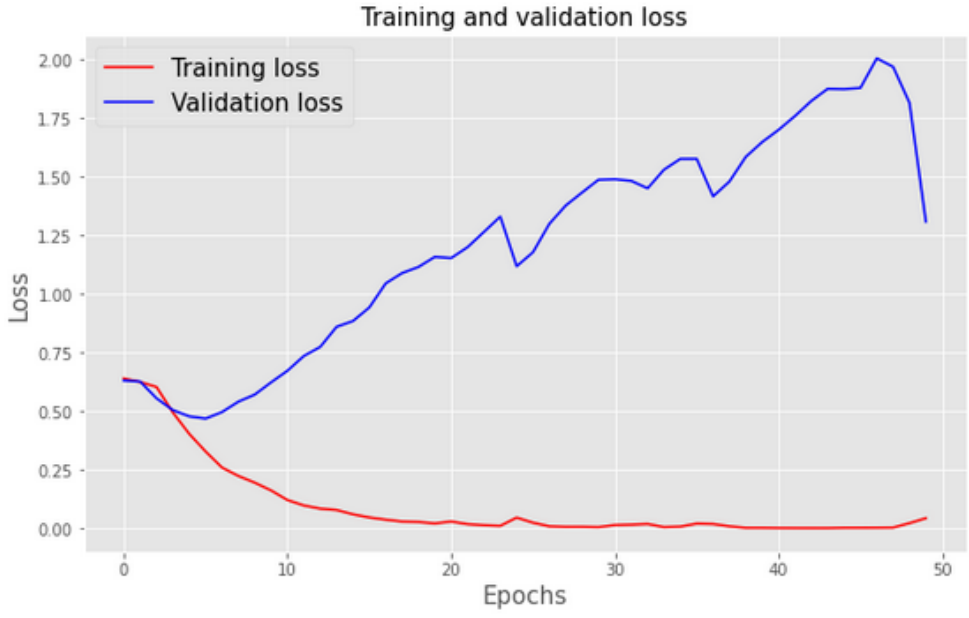


Figure 4.28: GRU Bidirectional Loss Graph

e) Multi-Dense LSTM Model- Figure 4.29 illustrates that when compared to LSTM, Multi-dense LSTM obtains good results, with a training accuracy of up to 91% and a validation accuracy of up to 84%. These figures compare favourably to the results obtained by LSTM. Figure 4.30 demonstrates a statistically significant difference between the training loss and the validation loss, which suggests that additional training epochs are required.

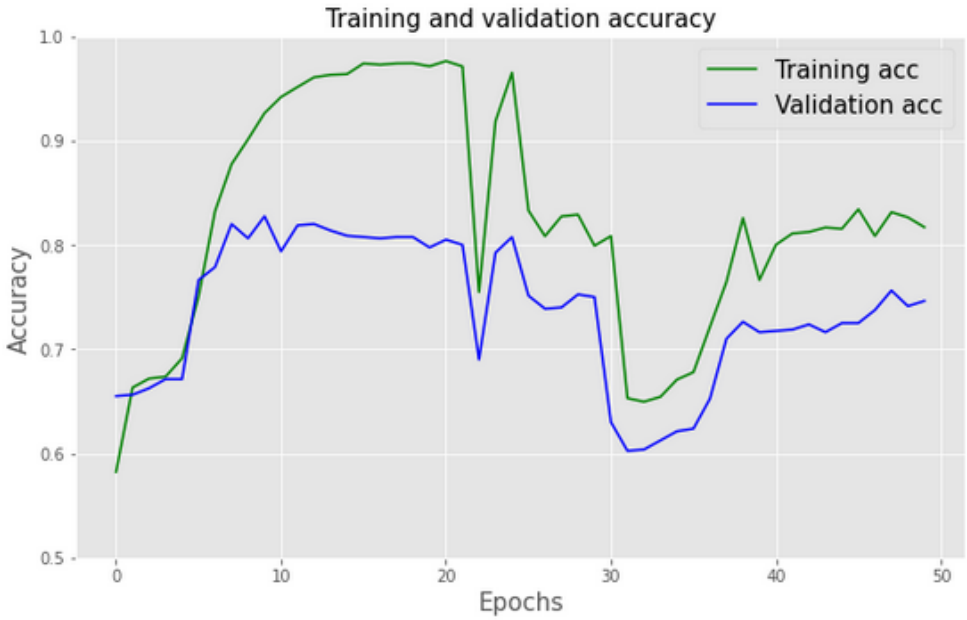


Figure 4.29: Multi-Dense LSTM Accuracy Graph

Figure 4.30: Multi-Dense LSTM Loss Graph

4.7 FRARBiLSTM:(With POS, LIWC, Bigram, Epoch=100)

The next phase involves the implementation of our Deep Learning Model FRARBiLSTM. This model will utilise POS, LIWC, Bigram, and Epoch=100 with Roberta on the AFINN-based dataset. These algorithms possess the ability to acquire knowledge and reach intelligent deductions autonomously. The data shown in Table 4.5 clearly shows that the RoBERTa algorithm's utilisation resulted in an accuracy enhancement.

Deep Learning Models	Accuracy without RoBERTa		Accuracy Enhanced with RoBERTa	
	Training	Validatio	Training	Validation
LSTM	90.71	80.62	93.71	91.69
FRARBiLSTM	**99.9**	**85.59**	**99.9**	**97.31**
Multi-dense LSTM	91.69	84.62	93.69	91.51
GRU	96.24	88.00	97.71	91.00
Bidirectional GRU	98.78	78.90	99.1	89.00

Table 4.5: FRARBiLSTM Performance Analysis

Thus, in this study, we introduced our initial Hybrid Model, FRARBiLSTM. This model outperforms other deep learning models and is regarded as superior, according to Figure 4.31. In today's context, crafting practical online assessments for a product or service is a significant concern that has become increasingly widespread and challenging to identify. This issue arises from effectively communicating the intended message to prospective clients. Subsequently, we compared the outcomes generated by the deep learning-based hybrid model FRARBiLSTM, which was previously introduced, and the outcomes generated by the ML-based and deep learning-based models, as depicted in Figure 4.34. The comparisons will be conducted based on the impacts generated by the hybrid model. Figure 4.32 and Figure 4.33 represent the interface we used to detect whether the review is Real or Fake. Table 4.6 compares the FRARBiLSTM model results with existing models discussed in the literature review section.

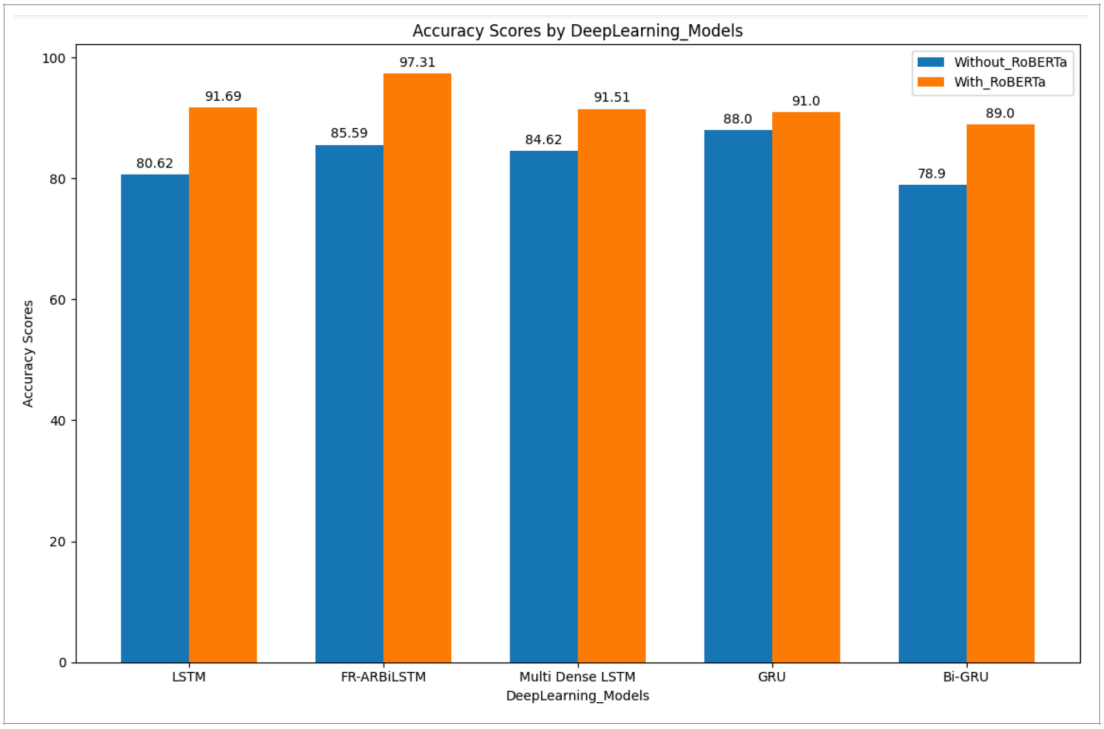


Figure 4.31: Accuracy Scores- FRARBiLSTM Vs.DeepLearning_Models

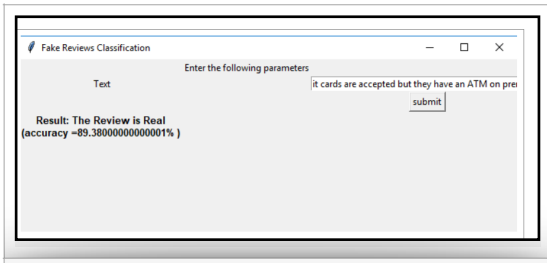


Figure 4.32 : Fake Review Classification interface-For Real Reviews

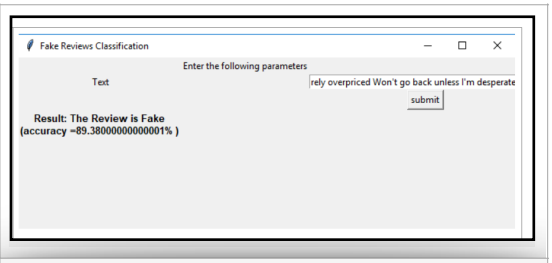


Figure 4.33 : Fake Review Classification interface-For Fake Reviews

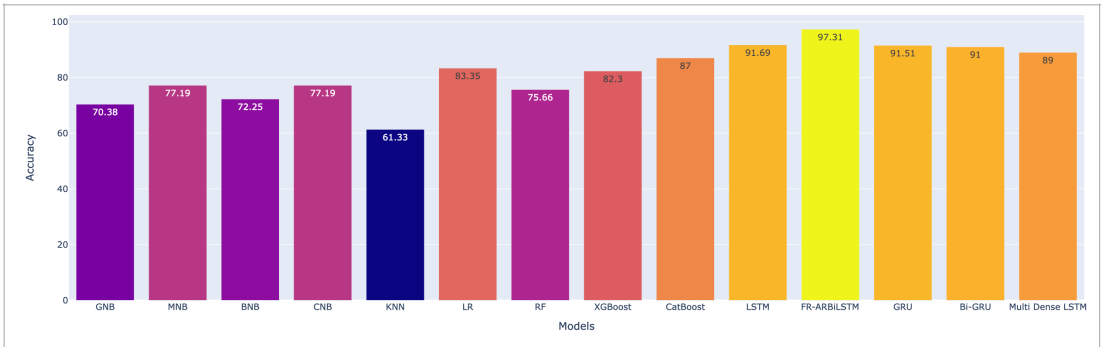


Figure 4.34 : FRARBiLSTM Vs. ML Vs. DeepLearning_Models

Models	Accuracy	
	Previous Study	This Study FRARBiLSTM + POS+LIWC+Bigram
Joni Salmine et al. [133]	96.64 (RoBERTa) *POS, Sentiments,Less Epochs*	**97.31** *Epochs=50,Epochs=100(Final)*
Deshai et al. [145]	93.8% (LSTM,CNN, RNN)	**97.31** *Epochs=50,Epochs=100(Final)*
Qadir et al. [144]	87.81%(ML Model+ BERT)	**97.31** *Epochs=50,Epochs=100(Final)*

Table 4.6: FRARBiLSTM Performance Vs. Existing Model

Under FRARBiLSTM, bidirectional LSTMs and GRU models capture text data's longest-term dependencies well. False reviewers may use terms or sentiment signals to hide their intent. LSTMs/GRUs can examine the entire review chain in both directions to find subtle trends and stylistic errors that may indicate a false review. Pre-trained transformer models like Roberta use word embeddings to express word meaning. Word context affects these embeddings for words. This lets the model analyse language and identify false review terms like excessive use of positive adjectives or generic assertions. Afinn uses a lexicon to analyse words' positive, negative, or neutral sentiments. Using sentiment analysis, the model can compare actual and false reviews' sentiment distribution.

Actual reviews communicate a broader spectrum of feelings, while fake ones may be unduly enthusiastic or negative to deceive the reader. These factors enhance a review's authenticity. While RNN models assess the sequential organisation and word usage, Afinn sentiment analysis adds an emotional tone. This allows the program to find patterns that separate legitimate reviews from fake ones.

Chapter 5

CONCLUSION AND FUTURE SCOPE

This study emphasised the importance of evaluations and their profound impact on virtually all facets of online content. Individuals decisions are significantly impacted by the credibility of online reviews and ratings, notwithstanding fraudulent content. As a result, detecting fake reviews is an active and evolving area of research that is still ongoing. This dataset was classified using a variety of different algorithms. The findings of this research reveal that our model FRARBiLSTM (Bi-LSTM+RoBERTa+AFINN) performs better than other ensemble classifiers (CatBoost+RoBERTa+AFINN) as well when compared with other deep learning models in the process of detecting fake reviews. Consequently, these algorithms can evaluate sentiment qualities and use just the review text to categorise reviews as legitimate or fraudulent.

Regarding the realisation of automatic fake reviewer detection on e-commerce platforms, our solution is a preliminary step in focusing on loyal consumers. Using the benchmark dataset, we compare FRARBiLSTM to multiple other baseline models to show its effectiveness.

Our model was superior to all of the different models and had the best accuracy of all of the models. In terms of overall performance, our model was successful. Our algorithm extracted the reviews' contextual, semantic, and sentiment information, resulting in outstanding detection results. We considered how sentiment knowledge might impact the detection of deceptive reviews, and as a result, we implemented BiLSTM to improve the model's overall performance. The experiments' conclusions showed that FRARBiLSTM succeeded in achieving outstanding results in detecting fraudulent reviews; moreover, the parameters that were taken into account needed to be more comprehensive. Suppose this is integrated with other pre-trained language models that consider more types of knowledge during the pre-training stage. However, future work may combine ensemble and deep learning models and employ more dynamic word embedding approaches as an integrated strategy.

In the future, we must construct multimodal data-based algorithms for detecting fake reviews. Data that includes text, visuals, and sound is said to be multimodal. Researchers can create more reliable and effective fake review detection models by having multimodal data that better captures the subtleties of human language and behaviour. One potential next step for studying fake review identification is to look into the application of adversarial learning. Adversarial learning is a method in which two models compete against one another. One model would be the fake review detection model, while the other would be the generator model used to fabricate the fake reviews. Instead of working together, the two models would fight against one another, with the fake review detection model looking for the false reviews and the generator model seeking to make reviews that can't be told apart from the actual ones. This approach helps enhance the accuracy of models used to identify fake reviews. Existing techniques for identifying deceptive reviews frequently need collecting extensive user information, such as names, emails, and IP addresses. Users' privacy may be compromised if this information is utilized to monitor their virtual movements. Privacy-preserving mechanisms, such as differential privacy, need to be developed by researchers so that fake review detection can be accomplished. It is crucial to check that models used to identify fake reviews are transparent and objective. This means that they can't favour some persons or goods over others. Researchers need to create tools to assess the reliability of review detection algorithms. This would be useful in preventing the models from being utilized in a discriminatory or harmful way.

The research results on authenticating fake reviews using our suggested model include the best features of several methods. By capturing text's long-term dependencies, bidirectional LSTMs or GRUs can spot stylistic flaws in fake evaluations. A pre-trained transformer model named Roberta uses contextual word embeddings to recognise ambiguities in language and identify expressions frequently linked to misleading material. Afinn sentiment analysis adds a last layer of insight into the review's emotional tone, which can differentiate between genuine and fake reviews. Combining these features, our proposed algorithm can learn intricate patterns distinguishing authentic and false reviews. Compared with more basic

models that depend on a single feature type, this could result in very accurate detection of bogus reviews. Roberta and similar pre-trained models can generalise to new data to a certain extent. Because of this, the model may do adequately on evaluations of other types of products or written in different styles, even though it needed to be trained on those particular variants.

Understanding these models can be challenging due to their various levels and intricate feature relationships. When dealing with massive datasets, the computational expense of training and running these models becomes even more apparent. This could be a problem in settings where resources are limited.

As a whole, the FRARBiLSTM model combines multiple feature analysis approaches in a way that shows promise for detecting fake reviews. One must know its limitations and possible biases to use it effectively and responsibly.

The approach can enhance the credibility of the online environment for both consumers and companies by detecting fake reviews. Consumers may improve their decision-making process by relying on authentic evaluations, while businesses can protect their brands against misleading information. By implementing effective optimisation techniques, the model has the potential to analyse reviews in real time. This capability enables platforms to identify and flag suspicious content, preventing its propagation promptly. This is essential for reducing the influence of fake reviews on consumer choices or a brand's reputation. The model's structure permits a certain degree of adjustment to various domains and writing styles. This can be advantageous for platforms that handle reviews across diverse product categories. The FRARBiLSTM model can be a valuable tool for real-time authentication of false reviews.

Roberta, a pre-trained model, can potentially acquire **biases** from the data it was trained on. When the model emphasises particular patterns linked to specific populations or writing styles, it can unjustly suppress valid evaluations. The complex characteristics of the FRARBiLSTM model create challenges in grasping the justification behind its decision-making process. The absence of transparency might be a problem since consumers may need to comprehend the logic behind flagging a review as fake.

The **scalability** of the FRARBiLSTM model for detecting false reviews offers benefits and constraints when applied to more extensive or more varied datasets. Roberta, a transformer model that has been pre-trained, is specifically built to process and manage vast quantities of textual input efficiently. As the training dataset size increases, the model's capacity to grasp complicated linguistic patterns and nuances might improve, leading to more excellent performance on larger datasets. The model architecture can be modified to accommodate larger datasets by augmenting the more hidden layers to enable the model.

Training and running this model on larger datasets might be computationally demanding. This necessitates robust hardware resources, which may not be easily accessible in all environments. More data sets are needed to ensure superior results. The quality and representativeness of the data are vital. If the larger dataset exhibits biases or lacks diversity in writing styles compared to the training data, the model's performance may not successfully expand. In general, the FRARBiLSTM model can handle larger datasets, but the availability of computational resources and the data quality become crucial factors to consider. The model's generalizability must be addressed for various datasets using strategies like domain adaptation.

Bibliography:

1. Rivaldo, Y., Kamanda, S. V., & Yusman, E. (2022). The Influence Of Brand Image, Promotion And Trust On Customer Loyalty At Bank BSI Nagoya Batam Branch. *Jurnal Mantik*, *6*(2), 2385-2392.
2. Román, S., Riquelme, I. P., & Iacobucci, D. (2023). Fake or credible? Antecedents and consequences of perceived credibility in exaggerated online reviews. *Journal of Business Research*, *156*, 113466.
3. Shivagangadhar, K., Sagar, H., Sathyan, S., & Vanipriya, C. H. (2015). Fraud detection in online reviews using machine learning techniques. *International Journal of Computational Engineering Research (IJCER)*, *5*(5), 52-56.
4. Martens, D., & Maalej, W. (2019). Towards understanding and detecting fake reviews in app stores. *Empirical Software Engineering*, *24*(6), 3316-3355.
5. Sun, C., Du, Q., & Tian, G. (2016). Exploiting product related review features for fake review detection. *Mathematical Problems in Engineering*, *2016*.
6. 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.
7. Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., & Shao, J. (2017, April). Bimodal distribution and co-bursting in review spam detection. In *Proceedings of the 26th international conference on world wide web* (pp. 1063-1072).
8. Li, Y., Lin, Y., Zhang, J., Li, J., & Zhao, L. (2015). Highlighting the fake reviews in review sequence with the suspicious contents and behaviours. *JOURNAL OF INFORMATION &COMPUTATIONAL SCIENCE*, *12*(4), 1615-1627.
9. Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., & Ghosh, R. (2013, August). Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 632-640).

10. Li, F. H., Huang, M., Yang, Y., & Zhu, X. (2011, June). Learning to identify review spam. In *Twenty-second international joint conference on artificial intelligence*.

11. Jindal, N., & Liu, B. (2007, May). Review spam detection. In Proceedings of the 16th international conference on World Wide Web (pp. 1189-1190).

12. Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. In *Proceedings of the international AAAI conference on web and social media* (Vol. 7, No. 1, pp. 175-184).

13. Jiang, M., Cui, P., & Faloutsos, C. (2016). Suspicious behavior detection: Current trends and future directions. *IEEE intelligent systems*, *31*(1), 31-39.

14. Algur, S. P., Ayachit, N. H., & Biradar, J. G. (2017). Exponential Distribution model for Review Spam Detection. *International Journal of Advanced Research in Computer Science*, *8*(3).

15. Rayana, S., & Akoglu, L. (2015, August). Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining* (pp. 985-994).

16. Mukherjee, A., Liu, B., & Glance, N. (2012, April). Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web* (pp. 191-200).

17. Lim, E. P., Nguyen, V. A., Jindal, N., Liu, B., & Lauw, H. W. (2010, October). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 939-948).

18. Ji, S. J., Zhang, Q., Li, J., Chiu, D. K., Xu, S., Yi, L., & Gong, M. (2020). A burst-based unsupervised method for detecting review spammer groups. *Information Sciences*, *536*, 454-469.

19. Rastogi, A., Mehrotra, M., & Ali, S. S. (2020). Effective opinion spam detection: A study on review metadata versus content. *Journal of Data and Information Science*, *5*(2), 76-110.

20. Lau, R. Y., Liao, S. Y., Kwok, R. C. W., Xu, K., Xia, Y., & Li, Y. (2012). Text mining and probabilistic language modeling for online review spam detection. *ACM Transactions on Management Information Systems (TMIS)*, *2*(4), 1-30.

21. Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, *311*, 18-38.

22. Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). What yelp fake review filter might be doing?. In *Proceedings of the international AAAI conference on web and social media* (Vol. 7, No. 1, pp. 409-418).

23. Akram, A. U., Khan, H. U., Iqbal, S., Iqbal, T., Munir, E. U., & Shafi, M. (2018). Finding rotten eggs: A review spam detection model using diverse feature sets.

24. Jindal, N., & Liu, B. (2008, February). Opinion spam and analysis. In Proceedings of the 2008 international conference on web search and data mining (pp. 219-230).

25. Ong, T., Mannino, M., & Gregg, D. (2014). Linguistic characteristics of shill reviews. *Electronic Commerce Research and Applications*, *13*(2), 69-78.

26. Rayson, P., Wilson, A., & Leech, G. (2002). Grammatical word class variation within the British National Corpus sampler. In *New frontiers of corpus research* (pp. 295-306). Brill.

27. Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2000). Longman grammar of spoken and written English.

28. Li, J., Ott, M., Cardie, C., & Hovy, E. (2014, June). 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)* (pp. 1566-1576).

29. Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *arXiv preprint arXiv:1107.4557*.

30. Shojaee, S., Murad, M. A. A., Azman, A. B., Sharef, N. M., & Nadali, S. (2013, December). Detecting deceptive reviews using lexical and syntactic features. In

2013 13th International Conference on Intellient Systems Design and Applications (pp. 53-58). IEEE.

31. Li, L., Ren, W., Qin, B., & Liu, T. (2015). Learning document representation for deceptive opinion spam detection. In Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data: 14th China National Conference, CCL 2015 and Third International Symposium, NLP-NABD 2015, Guangzhou, China, November 13-14, 2015, Proceedings 14 (pp. 393-404). Springer International Publishing.

32. Kim, S., Chang, H., Lee, S., Yu, M., & Kang, J. (2015, October). Deep semantic frame-based deceptive opinion spam analysis. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (pp. 1131-1140).

33. Jindal, N., & Liu, B. (2008, February). Opinion spam and analysis. In Proceedings of the 2008 international conference on web search and data mining (pp. 219-230).

34. Lin, Y., Zhu, T., Wang, X., Zhang, J., & Zhou, A. (2014, April). Towards online review spam detection. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 341-342).

35. Hernández-Castañeda, Á., Calvo, H., Gelbukh, A., & Flores, J. J. G. (2017). Cross-domain deception detection using support vector networks. *Soft Computing*, *21*, 585-595.

36. Sedighi, Z., Ebrahimpour-Komleh, H., & Bagheri, A. (2017, December). RLOSD: Representation learning based opinion spam detection. In *2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS)* (pp. 74-80). IEEE.

37. Khurshid, F., Zhu, Y., Yohannese, C. W., & Iqbal, M. (2017, November). Recital of supervised learning on review spam detection: An empirical analysis. In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)* (pp. 1-6). IEEE.

38. Khurshid, F., Zhu, Y., Hu, J., Ahmad, M., & Ahmad, M. (2022). Battering Review Spam Through Ensemble Learning in Imbalanced Datasets. *The Computer Journal*, *65*(7), 1666-1678.

39. Singhal, R., & Kashef, R. (2023). A Weighted Stacking Ensemble Model With Sampling for Fake Reviews Detection. *IEEE Transactions on Computational Social Systems*.

40. Deshai, N., & Rao, B. B. (2022). A detection of unfairness online reviews using deep learning. *J Theor Appl Inf Technol*, *100*(13), 4738-4779.

41. Khurshid, F., Zhu, Y., Xu, Z., Ahmad, M., & Ahmad, M. (2019). Enactment of ensemble learning for review spam detection on selected features. *International Journal of Computational Intelligence Systems*, *12*(1), 387-394.

42. Cardoso, E. F., Silva, R. M., & Almeida, T. A. (2018). Towards automatic filtering of fake reviews. *Neurocomputing*, *309*, 106-116.

43. Sánchez-Junquera, J., Villaseñor-Pineda, L., Montes-y-Gómez, M., & Rosso, P. (2018). Character N-grams for detecting deceptive controversial opinions. In Experimental IR Meets Multilinguality, Multimodality, and Interaction: 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings 9 (pp. 135-140). Springer International Publishing.

44. Mani, S., Kumari, S., Jain, A., & Kumar, P. (2018). Spam review detection using ensemble machine learning. In Machine Learning and Data Mining in Pattern Recognition: 14th International Conference, MLDM 2018, New York, NY, USA, July 15-19, 2018, Proceedings, Part II 14 (pp. 198-209). Springer International Publishing.

45. Nilizadeh, S., Aghakhani, H., Gustafson, E., Kruegel, C., & Vigna, G. (2019, May). Think outside the dataset: Finding fraudulent reviews using cross-dataset analysis. In *The World Wide Web Conference* (pp. 3108-3115).

46. Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., & Shao, J. (2017, April). Bimodal distribution and co-bursting in review spam detection. In

Proceedings of the 26th international conference on world wide web (pp. 1063-1072).

47. Sánchez-Junquera, J., Villasenor-Pineda, L., Montes-y-Gómez, M., Rosso, P., & Stamatatos, E. (2020). Masking domain-specific information for cross-domain deception detection. *Pattern Recognition Letters*, *135*, 122-130.

48. Mohawesh, R., Tran, S., Ollington, R., & Xu, S. (2021). Analysis of concept drift in fake reviews detection. *Expert Systems with Applications*, *169*, 114318.

49. Shan, G., Zhou, L., & Zhang, D. (2021). From conflicts and confusion to doubts: Examining review inconsistency for fake review detection. *Decision Support Systems*, *144*, 113513.

50. Yao, J., Zheng, Y., & Jiang, H. (2021). An ensemble model for fake online review detection based on data resampling, feature pruning, and parameter optimization. *Ieee Access*, *9*, 16914-16927.

51. Narayan, R., Rout, J. K., & Jena, S. K. (2018). Review spam detection using opinion mining. In Progress in Intelligent Computing Techniques: Theory, Practice, and Applications: Proceedings of ICACNI 2016, Volume 2 (pp. 273-279). Springer Singapore.

52. Siagian, A. H. A. M., & Aritsugi, M. (2020). Robustness of word and character n-gram combinations in detecting deceptive and truthful opinions. *Journal of Data and Information Quality (JDIQ)*, *12*(1), 1-24.

53. Martinez-Torres, M. D. R., & Toral, S. L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, *75*, 393-403.

54. Taneja, H., & Kaur, S. (2021). An ensemble classification model for fake feedback detection using proposed labeled CloudArmor dataset. *Computers & Electrical Engineering*, *93*, 107217.

55. 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.

56. Aiyar, S., & Shetty, N. P. (2018). N-gram assisted youtube spam comment detection. *Procedia computer science*, *132*, 174-182.

57. Nejad, S. J., Ahmadi-Abkenari, F., & Bayat, P. (2020, October). Opinion spam detection based on supervised sentiment analysis approach. In *2020 10th international conference on computer and knowledge engineering (ICCKE)* (pp. 209-214). IEEE.

58. Wang, X., Zhang, X., Jiang, C., & Liu, H. (2018, May). Identification of fake reviews using semantic and behavioral features. In *2018 4th International Conference on Information Management (ICIM)* (pp. 92-97). IEEE.

59. Noekhah, S., binti Salim, N., & Zakaria, N. H. (2020). Opinion spam detection: Using multi-iterative graph-based model. *Information processing & management*, *57*(1), 102140.

60. Mewada, A., & Dewang, R. K. (2023). A comprehensive survey of various methods in opinion spam detection. *Multimedia Tools and Applications*, *82*(9), 13199-13239.

61. Gao, Y., Gong, M., Xie, Y., & Qin, A. K. (2020). An attention-based unsupervised adversarial model for movie review spam detection. *IEEE transactions on multimedia*, *23*, 784-796.

62. Wang, H., Li, Z., Huang, J., Hui, P., Liu, W., Hu, T., & Chen, G. (2021, January). Collaboration based multi-label propagation for fraud detection. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (pp. 2477-2483).

63. Ennaouri, M., & Zellou, A. (2022). Fake Reviews Detection through Machine learning Algorithms: A Systematic Literature Review.

64. Hassan, R., & Islam, M. R. (2019, February). Detection of fake online reviews using semi-supervised and supervised learning. In *2019 International conference on electrical, computer and communication engineering (ECCE)* (pp. 1-5). IEEE.

65. Zhang, L., He, G., Cao, J., Zhu, H., & Xu, B. (2018). Spotting review spammer groups: A cosine pattern and network based method. *Concurrency and Computation: Practice and Experience*, *30*(20), e4686.

66. Li, L., Ren, W., Qin, B., & Liu, T. (2015). Learning document representation for deceptive opinion spam detection. In Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data: 14th China National Conference, CCL 2015 and Third International Symposium, NLP-NABD 2015, Guangzhou, China, November 13-14, 2015, Proceedings 14 (pp. 393-404). Springer International Publishing.

67. Zhao, S., Xu, Z., Liu, L., Guo, M., & Yun, J. (2018). Towards accurate deceptive opinions detection based on word order-preserving CNN. *Mathematical Problems in Engineering*, *2018*.

68. Wang, X., Liu, K., & Zhao, J. (2018). Detecting deceptive review spam via attention-based neural networks. In Natural Language Processing and Chinese Computing: 6th CCF International Conference, NLPCC 2017, Dalian, China, November 8–12, 2017, Proceedings 6 (pp. 866-876). Springer International Publishing.

69. Wang, X., Liu, K., & Zhao, J. (2017, July). Handling cold-start problem in review spam detection by jointly embedding texts and behaviors. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 366-376).

70. Zhang, W., Du, Y., Yoshida, T., & Wang, Q. (2018). DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network. *Information Processing & Management*, *54*(4), 576-592.

71. Li, Q., Wu, Q., Zhu, C., Zhang, J., & Zhao, W. (2019, July). An inferable representation learning for fraud review detection with cold-start problem. In *2019 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

72. Li, Q., Wu, Q., Zhu, C., Zhang, J., & Zhao, W. (2019). Unsupervised user behavior representation for fraud review detection with cold-start problem. In *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia*

Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part I 23 (pp. 222-236). Springer International Publishing.

73. Ren, Y., & Zhang, Y. (2016, December). Deceptive opinion spam detection using neural network. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 140-150).

74. Wang, C. C., Day, M. Y., Chen, C. C., & Liou, J. W. (2018, June). 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* (pp. 16-20).

75. Liu, W., Jing, W., & Li, Y. (2020). Incorporating feature representation into BiLSTM for deceptive review detection. *Computing*, *102*, 701-715.

76. Jain, N., Kumar, A., Singh, S., Singh, C., & Tripathi, S. (2019). Deceptive reviews detection using deep learning techniques. In Natural Language Processing and Information Systems: 24th International Conference on Applications of Natural Language to Information Systems, NLDB 2019, Salford, UK, June 26–28, 2019, Proceedings 24 (pp. 79-91). Springer International Publishing.

77. Zeng, Z. Y., Lin, J. J., Chen, M. S., Chen, M. H., Lan, Y. Q., & Liu, J. L. (2019). A review structure based ensemble model for deceptive review spam. *Information*, *10*(7), 243.

78. Dhamani, N., Azunre, P., Gleason, J. L., Corcoran, C., Honke, G., Kramer, S., & Morgan, J. (2019). Using deep networks and transfer learning to address disinformation. *arXiv preprint arXiv:1905.10412*.

79. Aghakhani, H., Machiry, A., Nilizadeh, S., Kruegel, C., & Vigna, G. (2018, May). Detecting deceptive reviews using generative adversarial networks. In *2018 IEEE Security and Privacy Workshops (SPW)* (pp. 89-95). IEEE.

80. You, Z., Qian, T., & Liu, B. (2018, August). An attribute enhanced domain adaptive model for cold-start spam review detection. In *Proceedings of the 27th international conference on computational linguistics* (pp. 1884-1895).

81. Tang, X., Qian, T., & You, Z. (2020). Generating behavior features for cold-start spam review detection with adversarial learning. *Information Sciences*, *526*, 274-288.

82. Wang, X., Liu, K., He, S., & Zhao, J. (2016, November). Learning to represent review with tensor decomposition for spam detection. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 866-875).

83. Wang, Y., Chan, S. C. F., Leong, H. V., Ngai, G., & Au, N. (2016). Multi-dimension reviewer credibility quantification across diverse travel communities. *Knowledge and information systems*, *49*, 1071-1096.

84. Heydari, A., Tavakoli, M., & Salim, N. (2016). Detection of fake opinions using time series. *Expert Systems with Applications*, *58*, 83-92.

85. Li, L., Qin, B., Ren, W., & Liu, T. (2017). Document representation and feature combination for deceptive spam review detection. *Neurocomputing*, *254*, 33-41.

86. Noekhah, S., Salim, N. B., & Zakaria, N. H. (2018). A novel model for opinion spam detection based on multi-iteration network structure. *Advanced Science Letters*, *24*(2), 1437-1442.

87. Yuan, C., Zhou, W., Ma, Q., Lv, S., Han, J., & Hu, S. (2019, November). Learning review representations from user and product level information for spam detection. In *2019 IEEE International Conference on Data Mining (ICDM)* (pp. 1444-1449). IEEE.

88. Cao, N., Ji, S., Chiu, D. K., He, M., & Sun, X. (2020). A deceptive review detection framework: Combination of coarse and fine-grained features. *Expert Systems with Applications*, *156*, 113465.

89. Fahfouh, A., Riffi, J., Mahraz, M. A., Yahyaouy, A., & Tairi, H. (2020). PV-DAE: A hybrid model for deceptive opinion spam based on neural network architectures. *Expert Systems with Applications*, *157*, 113517.

90. Guo, Z., Tang, L., Guo, T., Yu, K., Alazab, M., & Shalaginov, A. (2021). Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace. *Future generation computer systems*, *117*, 205-218.

91. Archchitha, K., & Charles, E. Y. A. (2019, September). Opinion spam detection in online reviews using neural networks. In *2019 19th International conference on advances in ICT for emerging regions (ICTer)* (Vol. 250, pp. 1-6). IEEE.

92. 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.

93. Budhi, G. S., Chiong, R., Wang, Z., & Dhakal, S. (2021). Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. *Electronic Commerce Research and Applications*, *47*, 101048.

94. 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).

95. Dadgar, S. M. H., Araghi, M. S., & Farahani, M. M. (2016, March). A novel text mining approach based on TF-IDF and Support Vector Machine for news classification. In *2016 IEEE International Conference on Engineering and Technology (ICETECH)* (pp. 112-116). IEEE.

96. Qu, S., Wang, S., & Zou, Y. (2008, November). Improvement of text feature selection method based on tfidf. In *2008 International Seminar on Future Information Technology and Management Engineering* (pp. 79-81). IEEE.

97. Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of sentimental reviews using machine learning techniques. *Procedia Computer Science*, *57*, 821-829.

98. Enríquez de Salamanca Ros, F., Troyano Jiménez, J. A., & López Solaz, T. (2016). An approach to the use of word embeddings in an opinion classification task. *Expert Systems with Applications, 66 (december 2016), 1-6.*

99. Soumya George, K., & Joseph, S. (2014). Text classification by augmenting bag of words (BOW) representation with co-occurrence feature. *IOSR Journal of Computer Engineering*, *16*(1), 34-38.

100. Wartena, C., Brussee, R., & Slakhorst, W. (2010, August). Keyword extraction using word co-occurrence. In *2010 workshops on database and expert systems applications* (pp. 54-58). IEEE.

101. Matsuo, Y., & Ishizuka, M. (2002). Keyword extraction from a document using word co-occurrence statistical information. *Transactions of the Japanese Society for Artificial Intelligence*, *17*(3), 217-223.

102. Kuang, Q., & Xu, X. (2010, August). Improvement and application of TF• IDF method based on text classification. In *2010 International Conference on Internet Technology and Applications* (pp. 1-4). IEEE.

103. Wang, L. H. (2014). An improved method of short text feature extraction based on words co-occurrence. *Applied Mechanics and Materials*, *519*, 842-845.

104. Albathan, M., Li, Y., & Algarni, A. (2012, December). Using patterns co-occurrence matrix for cleaning closed sequential patterns for text mining. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 201-205). IEEE.

105. Lott, B. (2012). Survey of keyword extraction techniques. *UNM Education*, *50*(10).

106. Kadhim, A. I., Cheah, Y. N., Ahamed, N. H., & Salman, L. A. (2014, December). Feature extraction for co-occurrence-based cosine similarity score of text documents. In *2014 IEEE student conference on research and development* (pp. 1-4). IEEE.

107. Zhou, S., Ling, T. W., Guan, J., Hu, J., & Zhou, A. (2003, March). Fast text classification: a training-corpus pruning based approach. In *Eighth International Conference on Database Systems for Advanced Applications, 2003.(DASFAA 2003). Proceedings.* (pp. 127-136). IEEE.

108. Tezgider, M., Yıldız, B., & Aydın, G. (2018, September). Improving word representation by tuning Word2Vec parameters with deep learning model. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (pp. 1-7). IEEE.

109.Ge, L., & Moh, T. S. (2017, December). Improving text classification with word embedding. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 1796-1805). IEEE.

110.Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*.

111.Rezaeinia, S. M., Ghodsi, A., & Rahmani, R. (2017). Improving the accuracy of pre-trained word embeddings for sentiment analysis. *arXiv preprint arXiv:1711.08609*.

112.ALRashdi, R., & O'Keefe, S. (2019). Deep learning and word embeddings for tweet classification for crisis response. *arXiv preprint arXiv:1903.11024*.

113.Kuyumcu, B., Aksakalli, C., & Delil, S. (2019, June). An automated new approach in fast text classification (fastText) A case study for Turkish text classification without pre-processing. In *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval* (pp. 1-4).

114.Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

115.Vora, P., Khara, M., & Kelkar, K. (2017). Classification of tweets based on emotions using word embedding and random forest classifiers. *International Journal of Computer Applications*, *178*(3), 1-7.

116.Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

117.Lilleberg, J., Zhu, Y., & Zhang, Y. (2015, July). Support vector machines and word2vec for text classification with semantic features. In *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)* (pp. 136-140). IEEE.

118.Stein, R. A., Jaques, P. A., & Valiati, J. F. (2019). An analysis of hierarchical text classification using word embeddings. *Information Sciences*, *471*, 216-232.

119. Raunak, V., Gupta, V., & Metze, F. (2019, August). Effective dimensionality reduction for word embeddings. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)* (pp. 235-243).

120. Elsaadawy, A., Torki, M., & Ei-Makky, N. (2018, December). A text classifier using weighted average word embedding. In *2018 International Japan-Africa Conference on Electronics, Communications and Computations (JAC-ECC)* (pp. 151-154). IEEE.

121. Ethayarajh, K. (2019). How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. *arXiv preprint arXiv:1909.00512*.

122. Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., & Okruszek, L. (2021). Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Research*, *304*, 114135.

123. Wu, X., Lv, S., Zang, L., Han, J., & Hu, S. (2019). Conditional bert contextual augmentation. In Computational Science–ICCS 2019: 19th International Conference, Faro, Portugal, June 12–14, 2019, Proceedings, Part IV 19 (pp. 84-95). Springer International Publishing.

124. Alsentzer, E., Murphy, J. R., Boag, W., Weng, W. H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. *arXiv preprint arXiv:1904.03323*.

125. Chang, W. C., Yu, H. F., Zhong, K., Yang, Y., & Dhillon, I. S. (2020, August). Taming pretrained transformers for extreme multi-label text classification. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 3163-3171).

126. Jwa, H., Oh, D., Park, K., Kang, J. M., & Lim, H. (2019). exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). *Applied Sciences*, *9*(19), 4062.

127. Maslennikova, E. (2019). ELMo Word Representations For News Protection. In *CLEF (Working Notes)*.

128. Chakraborty, R., Elhence, A., & Arora, K. (2019, September). Sparse Victory–A Large Scale Systematic Comparison of count-based and prediction-based vectorizers for text classification. In *Proceedings of the international conference on recent advances in natural language processing (RANLP 2019)* (pp. 188-197).

129. Manoharan, D. S. (2019). A smart image processing algorithm for text recognition, information extraction and vocalization for the visually challenged. *Journal of Innovative Image Processing*, *1*(1), 31-38.

130. Schwartz, A. (2020). Combining word embeddings for binary classification tasks.

131. Büyüköz, B., Hürriyetoğlu, A., & Özgür, A. (2020, May). Analyzing ELMo and DistilBERT on socio-political news classification. In *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020* (pp. 9-18).

132. Wang, C., Nulty, P., & Lillis, D. (2020, December). A comparative study on word embeddings in deep learning for text classification. In *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval* (pp. 37-46).

133. 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.

134. Worsham, J., & Kalita, J. (2020). Multi-task learning for natural language processing in the 2020s: where are we going?. *Pattern Recognition Letters*, *136*, 120-126.

135. Hou, Z., & Kung, S. Y. (2022). Semi-Supervised Few-Shot Learning from A Dependency-Discriminant Perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2817-2825).

136. Sun, Y., Wang, S., Li, Y., Feng, S., Tian, H., Wu, H., & Wang, H. (2020, April). Ernie 2.0: A continual pre-training framework for language understanding. In

Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 05, pp. 8968-8975).

137. Kuang, Y., Xu, H., Jiang, R., & Liu, Z. (2022, December). GTMS: A Gated Linear Unit Based Trust Management System for Internet of Vehicles Using Blockchain Technology. In *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 28-35). IEEE.

138. Jang, B., Kim, M., Harerimana, G., Kang, S. U., & Kim, J. W. (2020). Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism. *Applied Sciences*, *10*(17), 5841.

139. Wang, C., Ding, X., Zhou, X., Wang, C., & Liu, T. (2021). Depth wise bidirectional LSTM for speech emotion recognition. IEEE Signal Processing Letters, 28, 114-118.

140. Abduljabbar, R. L., Dia, H., & Tsai, P. W. (2021). Unidirectional and bidirectional LSTM models for short-term traffic prediction. *Journal of Advanced Transportation*, *2021*, 1-16.

141. Gao, Z., Li, Z., Luo, J., & Li, X. (2022). Short text aspect-based sentiment analysis based on CNN+ BiGRU. *Applied Sciences*, *12*(5), 2707.

142. Potamias, R. A., Siolas, G., & Stafylopatis, A. G. (2020). A transformer-based approach to irony and sarcasm detection. *Neural Computing and Applications*, *32*, 17309-17320.

143. https://jpt.spe.org/hybrid-machine-learning-explained-nontechnical-terms

144. Mir, A. Q., Khan, F. Y., & Chishti, M. A. (2023). Online Fake Review Detection Using Supervised Machine Learning And Bert Model. *arXiv preprint arXiv:2301.03225*.

145. Deshai, N., & Rao, B. B. (2022). Deep learning hybrid approaches to detect fake reviews and ratings. *Journal of Scientific & Industrial Research*, *82*(1), 120-127.

146. Du, J., Rong, J., Michalska, S., Wang, H., & Zhang, Y. (2019). Feature selection for helpfulness prediction of online product reviews: An empirical study. *PloS one*, *14*(12), e0226902.

147.Li, A., Liu, Z., Zhang, X., & Wang, W. (2019). FakeReviewGAN: A generative adversarial network for detecting fake reviews. *arXiv preprint arXiv:1901.08129*.

148.Wang, Y., Liu, Z., & Zhang, X. (2019). FakeReviewForest: A random forest based fake review detection model. *arXiv preprint arXiv:1901.07527*.

149.Chen, Y., Zhang, X., & Liu, Z. (2018). FakeReviewSVM: A support vector machine based fake review detection model. *arXiv preprint arXiv:1803.07840*.

150.Zhang, X., Liu, Z., & Wang, Y. (2017). FakeReviewLSTM: A long short-term memory neural network for detecting fake reviews. *arXiv preprint arXiv:1703.01473*.

151.Yang, Y., Chen, S., Wang, W., Zhang, X., & Liu, Z. (2022). FakeReviewBERT: A pre-trained language model for detecting fake reviews. *arXiv preprint arXiv:2201.04015*. https://arxiv.org/abs/2201.04015.

152.Chen, Y., Zhang, X., & Liu, Z. (2021). FakeReviewGAN-BERT: A generative adversarial network and BERT based fake review detection model. *arXiv preprint arXiv:2102.03685*.

153.Zhang, X., Liu, Z., & Wang, Y. (2020). FakeReviewGraph: A graph neural network based fake review detection model. arXiv preprint arXiv:2002.07334.

154.Wang, Y., Liu, Z., & Zhang, X. (2020). FakeReview-T5: A transfer learning approach for fake review detection. arXiv preprint arXiv:2002.08083.

155.Zhang, X., Liu, Z., & Wang, Y. (2022). FakeReview-RoBERTa: A robust fake review detection model based on RoBERTa. arXiv preprint arXiv:2201.00741.Zhang, X., Liu, Z., Wang, Y., & Liu, J. (2022).

156.Zhang, X., Wang, W., Zhang, S., Liu, Z., & Jindal, N. (2022). FakeReview-DPR: A dual-path representation learning approach for fake review detection. arXiv preprint arXiv:2203.00078.

157.Wang, Z., Zhang, J., Li, H., & Liu, L. (2017). Deepfakedet: A deep learning framework for detecting deepfakes. IEEE Transactions on Information Forensics and Security, 12(12), 2841-2853.

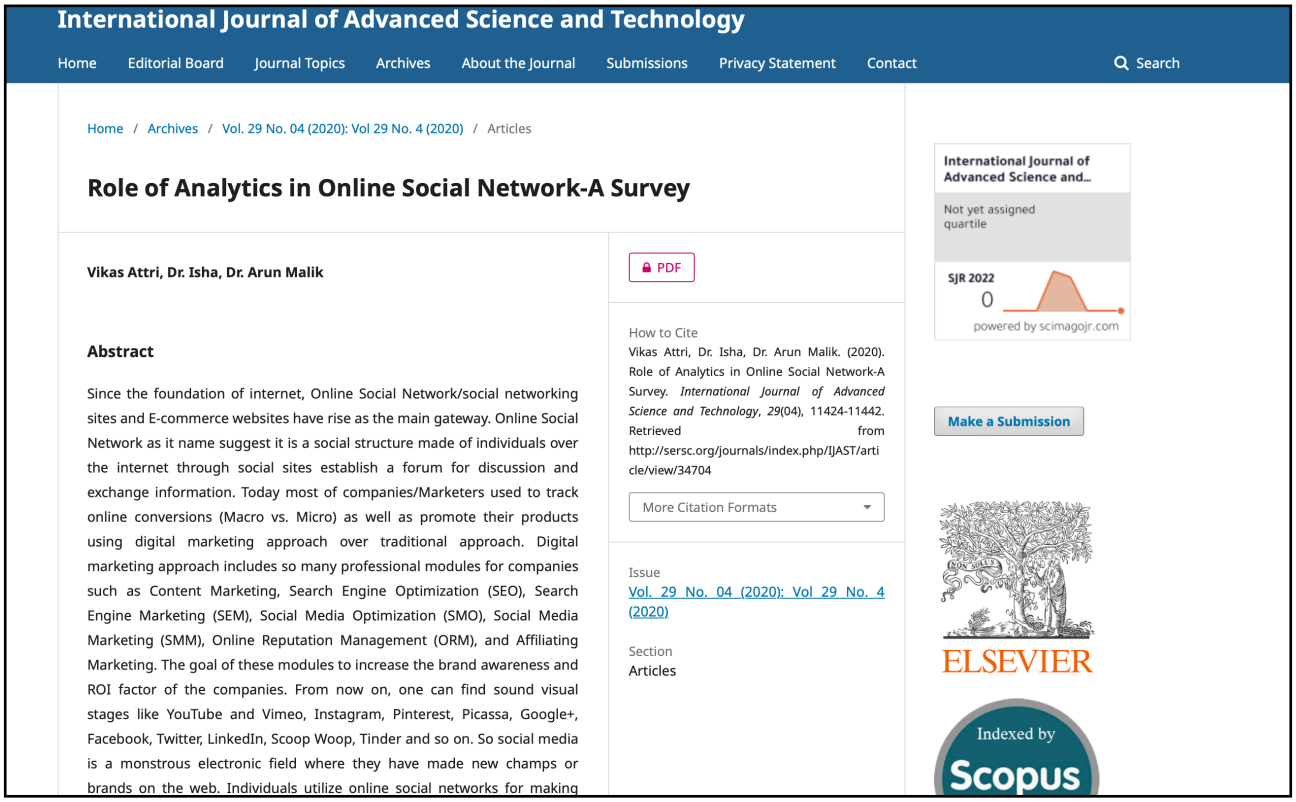
158.Liu, Y., Gao, J., Wang, C., & Liu, L. (2018). Reviewtrust: A trust-aware fake review detection model. IEEE Access, 6, 16786-16795.

159. Verma, V., Sharma, A., & Singh, S. (2019). FakingIt: A lexicon based approach for fake review detection. *Information Sciences*, 481, 286-300.

160. Zhang, J., Liu, L., & Chen, H. (2020). Fakereviewerdet: A novel fake reviewer detection model based on deep learning. *Information Sciences*, 532, 459-473.

Index

Appendix -I Review paper in Scopus indexed Journal,International Journal of Advanced Science and Technology (Role of Analytics in Online Social Network-A Survey)



International Journal of Advanced Science and Technology

Home Editorial Board Journal Topics Archives About the Journal Submissions Privacy Statement Contact Search

Home / Archives / Vol. 29 No. 04 (2020): Vol 29 No. 4 (2020) / Articles

Role of Analytics in Online Social Network-A Survey

Vikas Attri, Dr. Isha, Dr. Arun Malik

PDF

Abstract

Since the foundation of internet, Online Social Network/social networking sites and E-commerce websites have rise as the main gateway. Online Social Network as it name suggest it is a social structure made of individuals over the internet through social sites establish a forum for discussion and exchange information. Today most of companies/Marketers used to track online conversions (Macro vs. Micro) as well as promote their products using digital marketing approach over traditional approach. Digital marketing approach includes so many professional modules for companies such as Content Marketing, Search Engine Optimization (SEO), Search Engine Marketing (SEM), Social Media Optimization (SMO), Social Media Marketing (SMM), Online Reputation Management (ORM), and Affiliating Marketing. The goal of these modules to increase the brand awareness and ROI factor of the companies. From now on, one can find sound visual stages like YouTube and Vimeo, Instagram, Pinterest, Picassa, Google+, Facebook, Twitter, LinkedIn, Scoop Woop, Tinder and so on. So social media is a monstrous electronic field where they have made new champs or brands on the web. Individuals utilize online social networks for making

How to Cite
Vikas Attri, Dr. Isha, Dr. Arun Malik. (2020). Role of Analytics in Online Social Network-A Survey. *International Journal of Advanced Science and Technology*, *29*(04), 11424-11442. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/34704

More Citation Formats

Issue
Vol. 29 No. 04 (2020): Vol 29 No. 4 (2020)

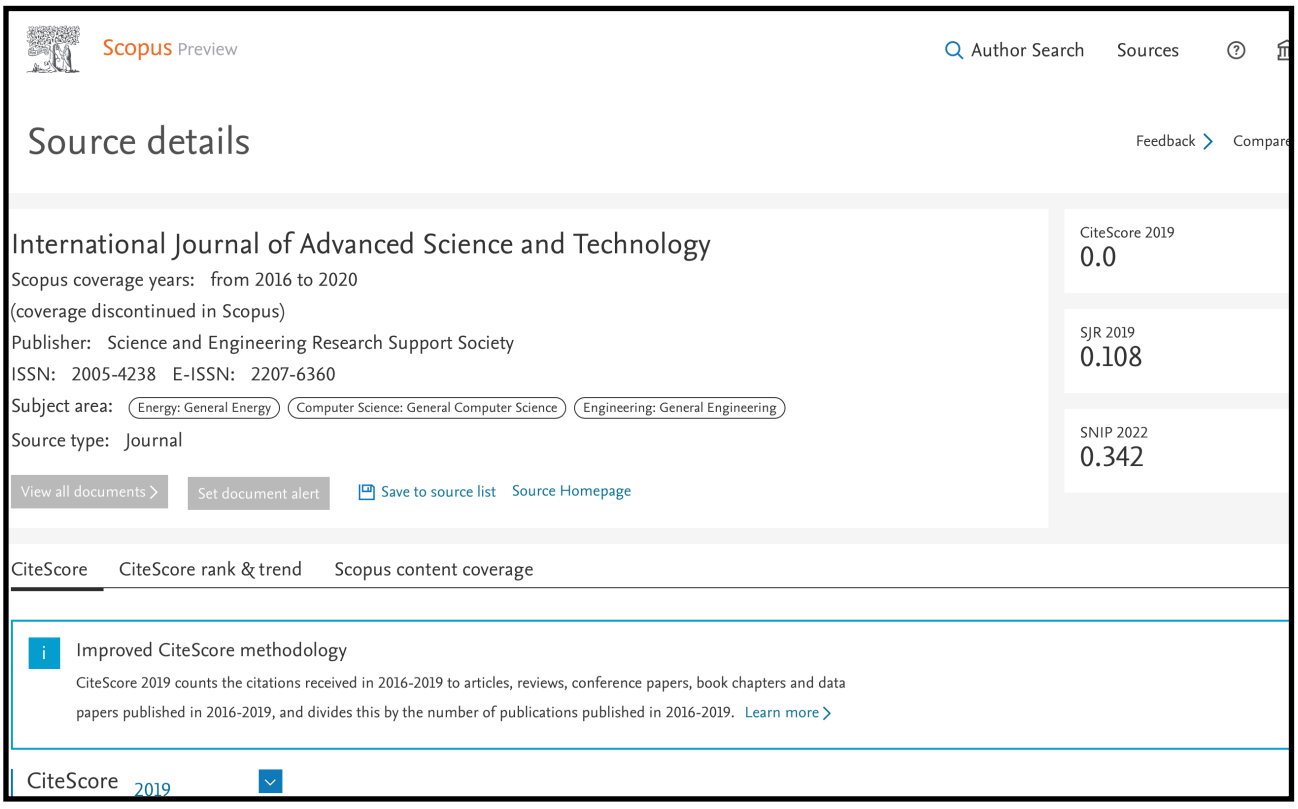
Section
Articles

International Journal of Advanced Science and...
Not yet assigned quartile
SJR 2022
0
powered by scimagojr.com

Make a Submission

ELSEVIER

Indexed by Scopus



Scopus Preview

Author Search Sources

Source details

Feedback Compare

International Journal of Advanced Science and Technology
Scopus coverage years: from 2016 to 2020
(coverage discontinued in Scopus)
Publisher: Science and Engineering Research Support Society
ISSN: 2005-4238 E-ISSN: 2207-6360
Subject area: Energy: General Energy | Computer Science: General Computer Science | Engineering: General Engineering
Source type: Journal

View all documents | Set document alert | Save to source list | Source Homepage

CiteScore 2019
0.0

SJR 2019
0.108

SNIP 2022
0.342

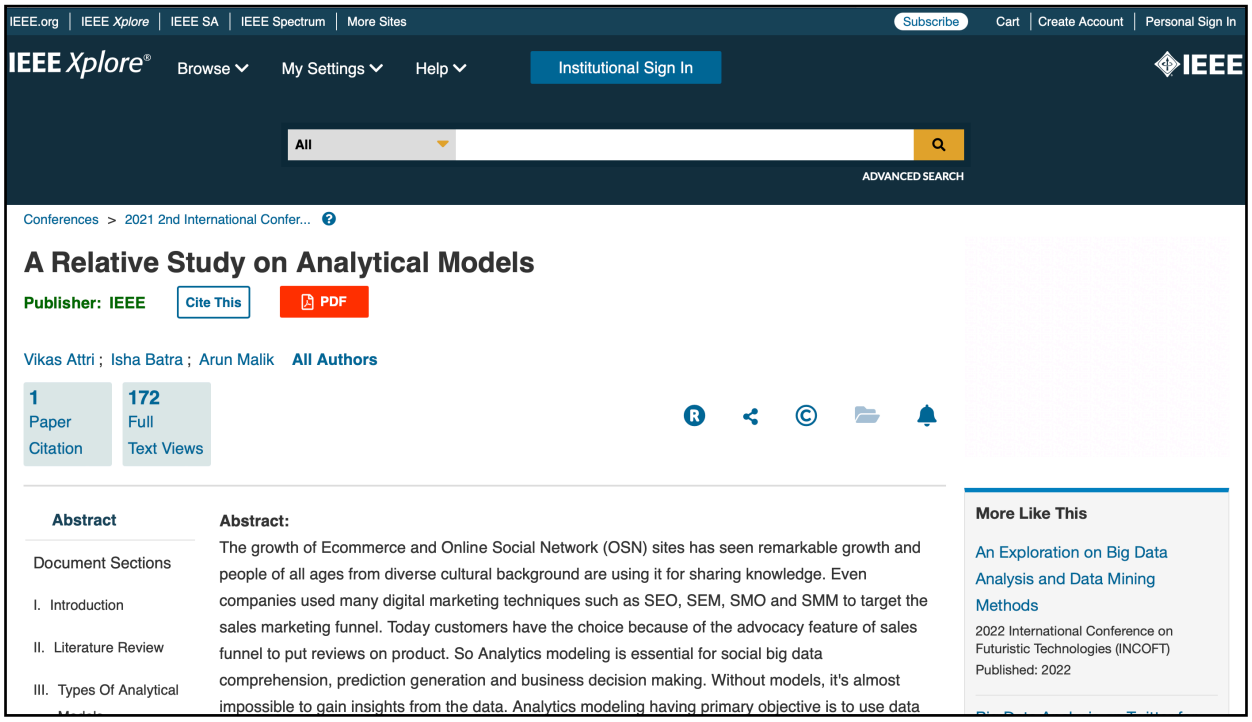
CiteScore | CiteScore rank & trend | Scopus content coverage

Improved CiteScore methodology
CiteScore 2019 counts the citations received in 2016-2019 to articles, reviews, conference papers, book chapters and data papers published in 2016-2019, and divides this by the number of publications published in 2016-2019. Learn more

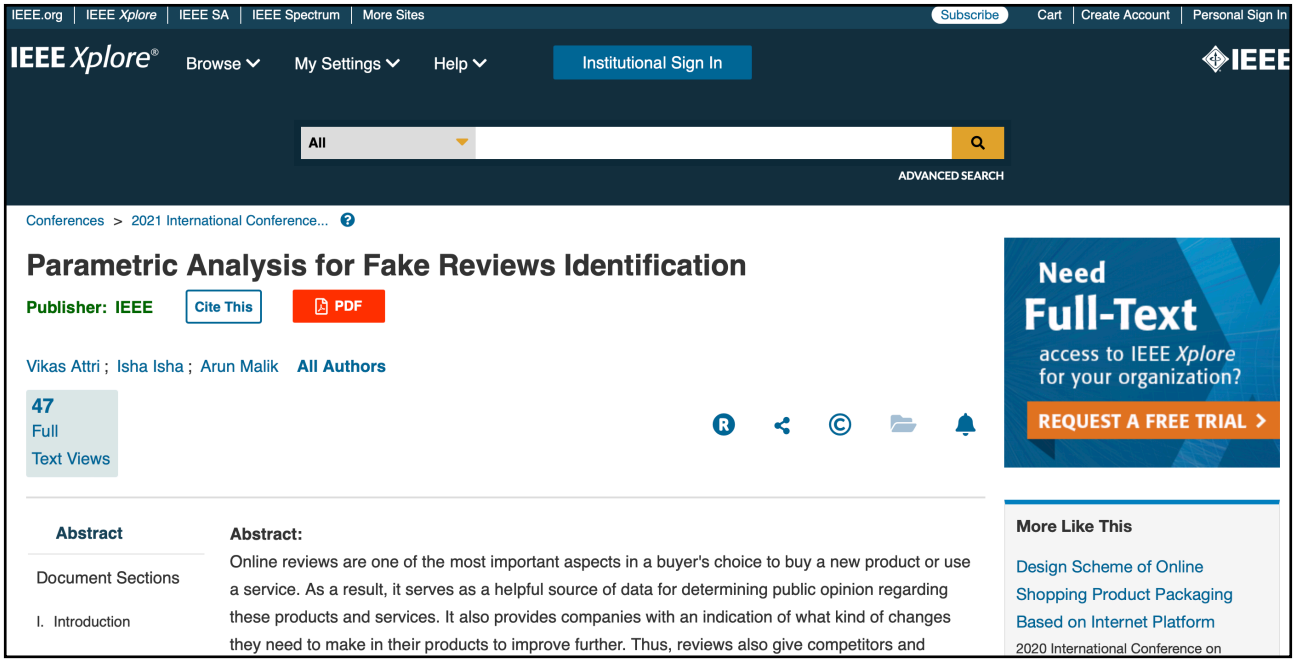
CiteScore 2019

Vol. 29 No. 04 (2020): Vol 29 No. 4 (2020)

Appendix -II Conference Paper presented in the 2021, 2nd International Conference on Intelligent Engineering and Management (ICIEM)(A Relative Study on Analytical Models)



Appendix -III Conference Paper presented in the 2021, International Conference on Computing Sciences (ICCS) (Parametric Analysis for Fake Reviews Identification)



IEEE.org | IEEE Xplore | IEEE SA | IEEE Spectrum | More Sites | Subscribe | Cart | Create Account | Personal Sign In

IEEE Xplore® Browse My Settings Help Institutional Sign In IEEE

All

ADVANCED SEARCH

Conferences > 2021 International Conference...

Parametric Analysis for Fake Reviews Identification

Publisher: IEEE | Cite This | PDF

Vikas Attri ; Isha Isha ; Arun Malik All Authors

47 Full Text Views

Abstract

Document Sections

I. Introduction

Abstract:

Online reviews are one of the most important aspects in a buyer's choice to buy a new product or use a service. As a result, it serves as a helpful source of data for determining public opinion regarding these products and services. It also provides companies with an indication of what kind of changes they need to make in their products to improve further. Thus, reviews also give competitors and

Need Full-Text access to IEEE Xplore for your organization?

REQUEST A FREE TRIAL >

More Like This

Design Scheme of Online Shopping Product Packaging Based on Internet Platform

2020 International Conference on

Certificate No. 236384

ICCS

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

PUBLISHED BY IEEE COMPUTER SOCIETY CONFERENCE PUBLISHING SERVICES

5th INTERNATIONAL CONFERENCE ON COMPUTING SCIENCES (ICCS-2021)

Certificate of Paper Presentation

This is to certify that Dr./Mr./Ms. Vikas Attri

of Lovely Professional University, Punjab has presented

paper entitled Parametric Analysis for Fake Reviews Identification

in 5th International Conference on Computing Sciences (ICCS) "Kathleen - 100" held on 4-5th December, 2021

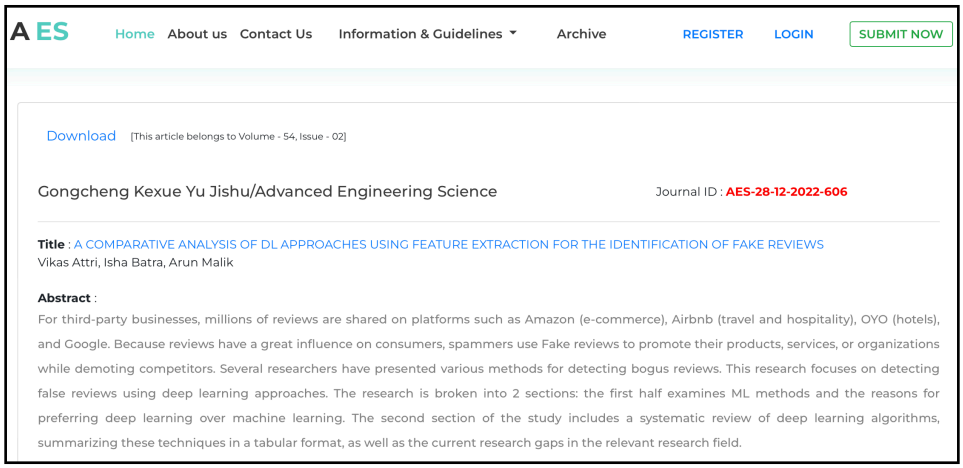
organized by Lovely Faculty of Technology and Sciences, Lovely Professional University, Punjab.

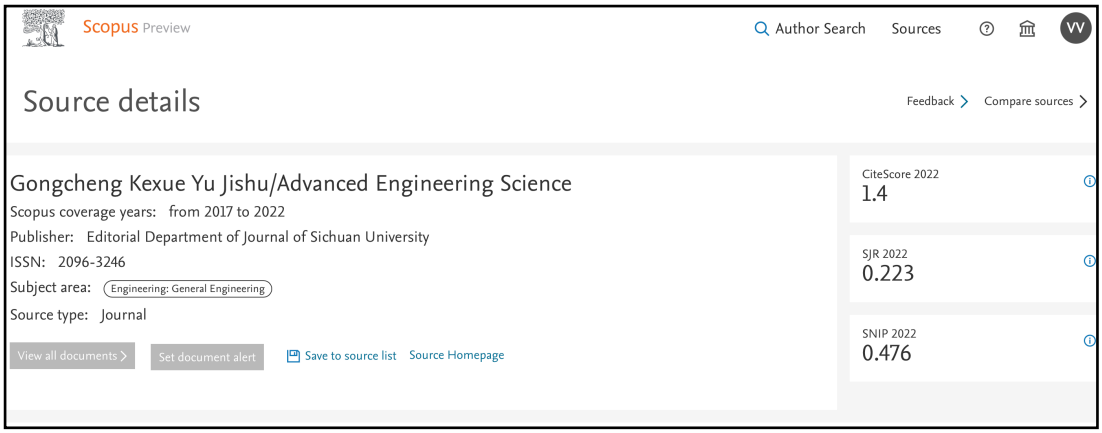
Date of Issue : XXXXX

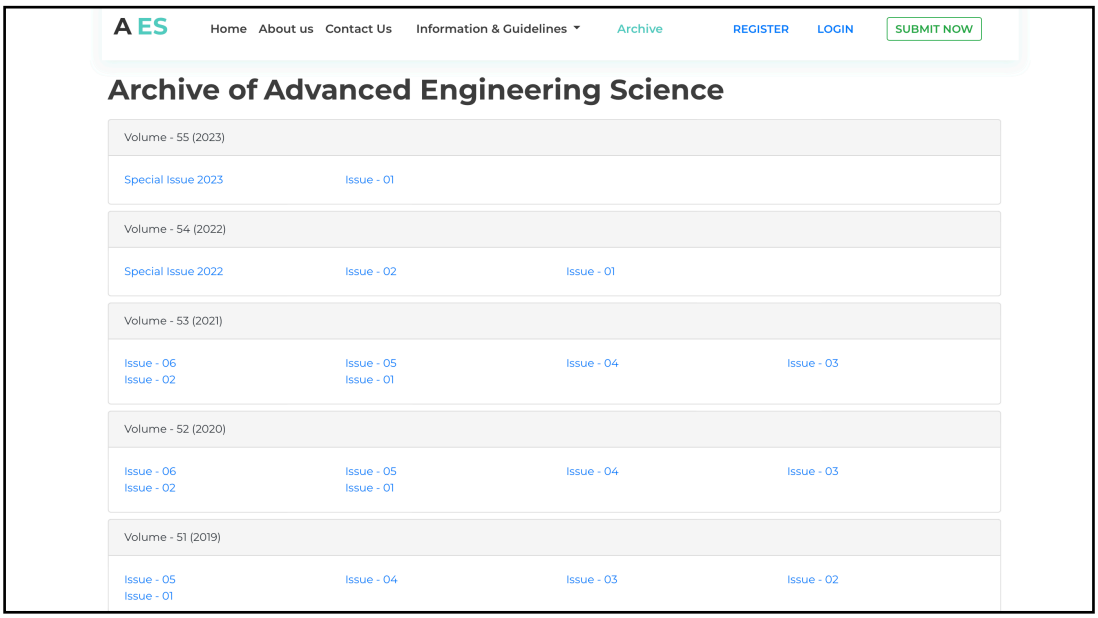
Place of Issue: Phagwara (India)

Prepared by (Administrative Officer-Records) | General Chair | Program Chair | Conference Chair

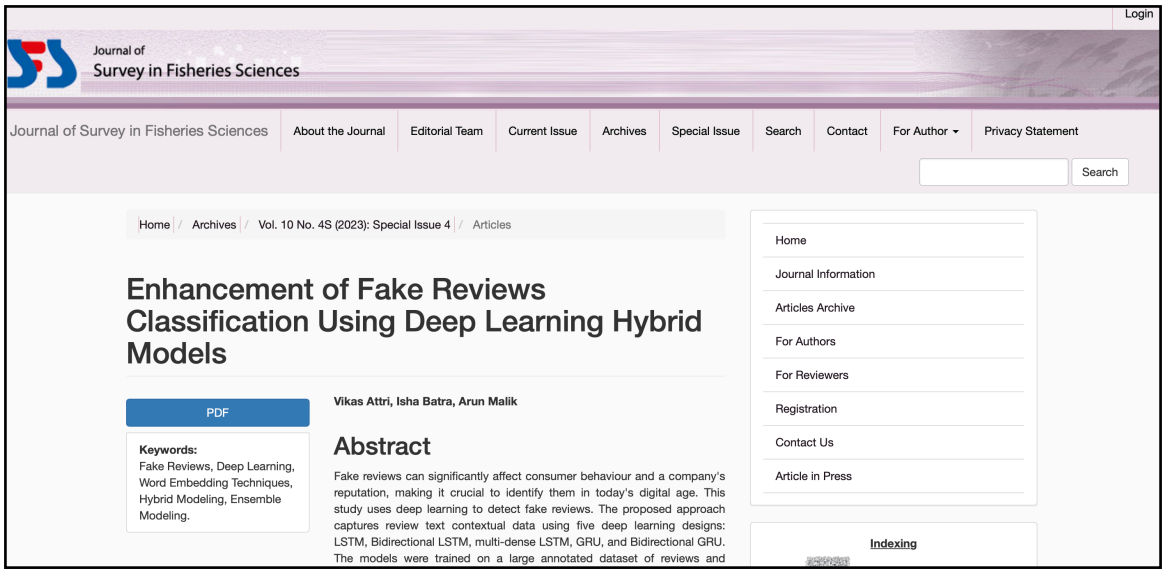
Appendix -IV Paper in Scopus indexed Journal, Gongcheng Kexue Yu Jishu/Advanced Engineering Science(A comparative analysis of dl approaches using feature extraction for the identification of fake reviews

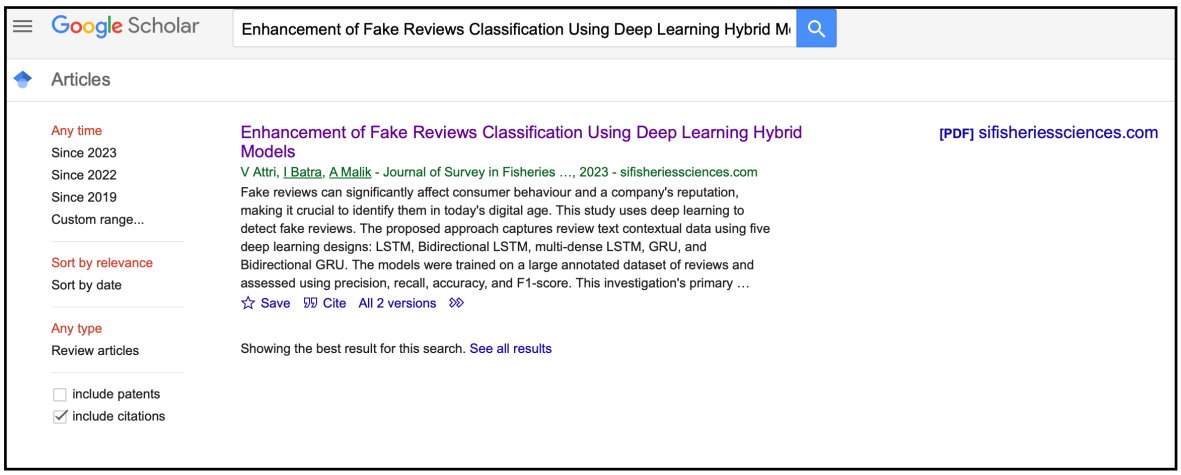






Appendix -V Paper in Web of Science indexed Journal, Journal of Survey in Fisheries Sciences (SFS) (Enhancement of Fake Reviews Classification Using Deep Learning Hybrid Models)





Appendix -VI Paper in Scopes indexed Journal, International Journal of Intelligent Systems and Applications in Engineering, FRARBiLSTM- A Novel Fake Review Authentication Model Using Afinn and Roberta

FRARBiLSTM-A Novel Fake Review Authentication Model Using Afinn and Roberta

Authors	Vikas Attri, Isha Batra, Arun Malik, Vipin Kumar
Publication date	2024/1/7
Journal	International Journal of Intelligent Systems and Applications in Engineering
Volume	12
Issue	10s
Pages	135-144

Article

FRARBiLSTM-A Novel Fake Review Authentication Model Using Afinn and Roberta

Attri, V., Batra, I., Malik, A., Kumar, V.

International Journal of Intelligent Systems and Applications in Engineering, 2024, 12(10s), pp. 135–144

Show abstract Related documents