INTEGRATING DEEP LEARNING AND NLP FOR EFFECTIVE IDENTIFICATION OF FAKE NEWS IN SOCIAL MEDIA CONTENT ANALYSIS

by

Karthik Bysani Sathyanarayana

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

JUNE, 2025

INTEGRATING DEEP LEARNING AND NLP FOR EFFECTIVE IDENTIFICATION OF FAKE NEWS IN SOCIAL MEDIA CONTENT ANALYSIS

by

Karthik Bysani Sathyanarayana

Supervised by

Dr. Ibrahim Menkeh Muafueshiangha

APPROVED BY

Anna Provodnikova, PhD

Dissertation chair

RECEIVED/APPROVED BY:

Admissions Director

Dedication

I dedicate this thesis to my beloved family, whose unwavering support made this journey possible.

To my **parents**, for their endless love, patience, and understanding despite my absence during challenging times.
To my **daughter**, whose innocence and smile inspired me to persevere even when I couldn't spend quality time with her.
To my **newborn son**, who brought boundless joy amidst the chaos, reminding me of the beauty of life.
To my **spouse**, for standing by my side through the demands of work, financial challenges, and countless sacrifices.

A special note of gratitude to my **mentor**, whose guidance, encouragement, and swift turnaround helped accelerate my DBA journey.

This achievement is as much yours as it is mine.

.

Acknowledgements

I would like to express my deepest gratitude to all those who supported me throughout the journey of pursuing my Doctorate of Business Administration.

First and foremost, my sincere thanks to my **supervisor/mentor [Dr. Ibrahim Menkeh Muafueshiangha]** for their invaluable guidance, encouragement, and prompt feedback. Your expertise, patience, and unwavering support played a crucial role in shaping this research and accelerating my DBA journey.

I am immensely grateful to my **family** — my **parents**, for their unconditional love and blessings; my **spouse**, for standing by me with patience and resilience; and my **daughter and newborn son**, whose smiles brought light during the most challenging moments. This journey would not have been possible without their sacrifices and constant encouragement, especially during difficult times with job challenges, financial constraints, and health concerns.

A special mention to my **friends, colleagues, and professional network** who provided their encouragement, shared knowledge, and acted as sounding boards during this academic journey.

Finally, I extend my heartfelt gratitude to the **institution [Swiss School of Business and Management]** and its faculty for providing me with the platform and resources to pursue this academic endeavor.

This journey has been a testament to the power of perseverance, faith, and the unwavering support of those around me.

ABSTRACT

INTEGRATING DEEP LEARNING AND NLP FOR EFFECTIVE IDENTIFICATION OF FAKE NEWS IN SOCIAL MEDIA CONTENT ANALYSIS

Karthik Bysani Sathyanarayana
2025

Dissertation Chair: Dr. Anna Provodnikova
Co-Chair: Dr. Ljiljana Kukec

Social media is becoming the primary news source globally, but the quick dissemination of false news on these platforms is a major global issue that affects social, political, and economic institutions as well as public perception by evoking feelings like contempt, surprise, and terror. Therefore, identifying and thwarting false news has become essential to maintaining the credibility of information found online. This study draws on Framing Theory and the Diffusion of Misinformation Theory to better understand how deceptive content spreads and is perceived on social platforms. A major limitation of prior research lies in its struggle to accurately capture deep contextual nuances, which this study addresses using the Roberta sequence classifier, known for its rich contextual embeddings. The research utilizes two widely recognized datasets—PolitiFact and Gossip Cop—and implements a robust preprocessing pipeline involving text normalization, tokenization, and class imbalance handling via random oversampling. The purpose of exploratory data analysis (EDA) is to uncover underlying patterns. Standard measures like as accuracy, precision, recall, F1-score, AUC, and confusion matrix are used to assess the refined Roberta model. Results show superior performance, achieving 93.55% accuracy on PolitiFact and 94.19% on Gossip Cop, outperforming models. This work presents an accurate, scalable solution for fake news detection, grounded in theoretical insights and demonstrating enhanced context comprehension and classification performance.

Keywords: *Social media, Fake news, Machine learning, Roberta, PolitiFact, and Gossip Cop.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviations	Full Form
HCI	Human-Computer Interactions
AI	Artificial Intelligence
DL	Deep Learning
ML	Machine Learning
NLP	Natural Language Processing
DBM	Deep Belief Networks
DNN	Deep Neural Network
RNN	Recurrent Neural Network
FFNN	Feed Forward Neural Network
CNN	Convolution Neural Networks
RBMs	Restricted Boltzmann Machines
LSTM	Long Short-Term Memory
LRN	Local Response Normalization
VGG	Visual Geometry Group
CEC	Constant Error Carousel
LLMs	Large Language Models
EDA	Exploratory Data Analysis
SSD	Solid-State Drive
ZSL	Zero-Shot Learning
MICE	Multiple Imputation Chain Equation
GRNN	Gated Recurrent Neural Network

DSAE	Deep Stacked Auto Encoder
FL	Federated Learning
MCS	Mobile Crowdsensing
FND	Fake News Detection
ABC	Adaptive Boosting Classifier
SVM	Support Vector Machine
SMOTE	Synthetic Minority Oversampling Technique
GRU	Gated Recurrent Unit
FNC	Fake News Challenges

CHAPTER I:

INTRODUCTION

1.1 Overview

The Internet is an important invention, and many people use it for different purposes. There are several problems in our digital age. Among them is fake news (FN). The dissemination of false information may damage a reputation of an individual or a company with relative ease. Propaganda against an individual, political party, or organization may be the source (Ahmed, Aljabouh and Praveen Kumar Donepudi, 2021). The proliferation, cheap cost, and ease of use of social media for news consumption have all contributed to its meteoric rise in popularity over the last decade. The rapid worldwide dissemination of "fake news," or purposely misleading articles, is a potential drawback of social media. Negative and far-reaching societal effects might result from social media FN. Many people are interested in studying how to spot FN on social media, which is a relatively new phenomenon. M. Khan et al. (2020) Fake news has been categorized statistically, and ML models employ these classes. Theoretical frameworks guide applications of Human-Computer Interactions (HCI) and ML models in particular (Series, 2021).

There is an immediate and vital need to think about ways to detect fake news early on because of how quickly it travels and changes on social media. This task becomes much more onerous considering how often platforms like social media evolve. Either content-based or social context-based approaches to early detection of fake news provide a substantial barrier (Aïmeur, Amri and Brassard, 2023). As smartphones become more widely used, people are able to access social media practically anywhere, at any time, unlike with conventional media. As a result, people are spending more and more time participating on these platforms. The purpose of fake news is to spread false information or propaganda. However, in order to conceal logical reactions, analysis, and comparison of data from other sources, it consistently plays on public emotions, promoting indignation and indignation. It can quickly result in politicized, biassed information that harms other people and conspiracy theories. As a result of its prevalence, subsidization, and convenience of one-click availability, the public depends on online news articles and social media as their principal sources of information and news (Balshetwar, Abilash and Dani Jermisha, 2023). A 2020 Norwegian poll found that misinformation on the coronavirus was most often spread via social media. Figure 1.1 shows how various types of media have been impacted by social media. Therefore, in order to effectively avoid risks and harm, early identification of counterfeit photographs on social media sites is vital (Sharma *et al.*, 2023).

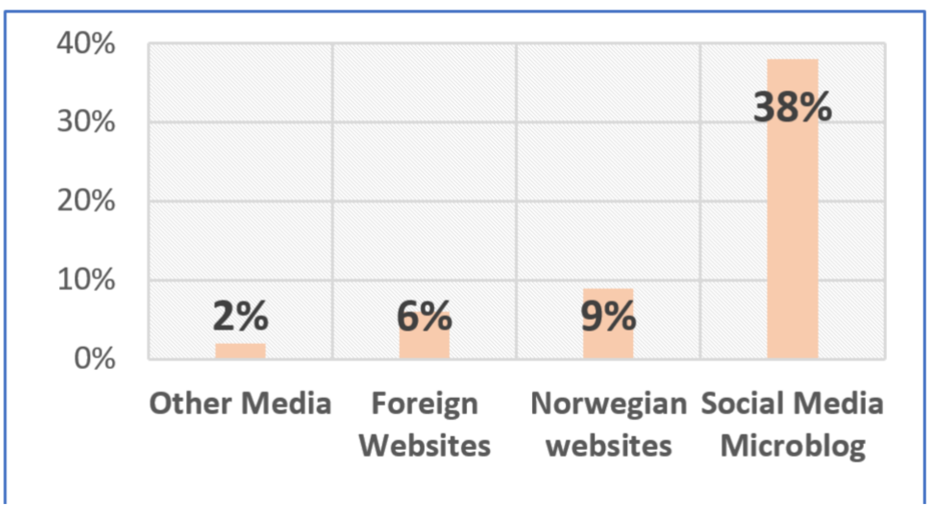

Figure 1.1: Social media is the highest contributor to fake news.

1.2 Definition and History

Articles that are obviously false and intended to mislead readers are called "fake news." We concentrate on political fake news stories, paying particular attention to the US presidential election of 2016. One of the commonly shared pieces by the now-defunct website denverguardian.com that belongs to a category of purposely fake news articles is

"Suicide seems to have been the cause of death for FBI agent implicated in Hillary email leaks." Also included are several things sourced from satirical websites, which, when seen in isolation on sites like Facebook or Twitter, may be mistaken for actual news. An example of this would be the July 2016 claim made by the now-defunct wtoe5news.com, which stated that Pope Francis had supported Donald Trump's presidential campaign. The "About" section of WTOE 5News revealed that it is "a fantasy news website." The majority of the stories on wtoe5news.com are either pure fiction or parody, yet the article did not carry this notice. More than a million people shared the story on Facebook, and some people who took part in our study (which you can read about below) really thought the headline was true.

Our definition excludes several closely related forms of fake news: 1) the occasional inaccuracy in reporting, such the false rumour that President Trump had taken a bustof Martin Luther King Jr. down by the Oval Office in the Whitehouse; 2) rumours that don't come by a specific news story;1 3) conspiracy theories (which are usually created by those who think there are true and are by definition hard to prove to be accurate or wrong);2 4) uncanny humour that most people would take seriously; 5) politicians who make false accusations; and 6) stories that are biassed or deceptive but not completely untrue (Gentzkow, Shapiro, and Stone 2016 define FN as "distortion," not "filtering")(Allcott and Gentzkow, 2017).

1.3 Social media as a news source

Earlier, media persons used to have their exclusive news sources, but with social media becoming one of the major sources of information this exclusivity is lost. argue that the onset of online and social media has enabled direct communication between newspersons and the public. In fact, the public can disseminate or hold back information on social platforms until the time want. Sources of news are not in the control of newspersons anymore. Major broadcasters, newspapers, and magazines are seen incorporating content generated on social media in their news stories. Twitter has become a platform where a politician, a celebrity, a corporate, or an ordinary citizen can voice their opinion and other users can receive updates and respond(Broersma and Graham, 2013).

It is possible for individuals to share diverse values and beliefs by giving them unrestricted access to a vast amount of knowledge. The majority of people are still unsure of the dangers and ramifications of this new resource. Fake news is one of these risks. FN is not verified, yet it seems realistic and polished enough that consumers may mistake it for true news (TN). A good illustration of how FN affects many parts of society is how it influences SM, which in turn affects the reactions of governments, organizations, and individuals to societal events. A large portion of FN is aimed towards a certain demographic in an effort to promote an ideology through the inculcation of strong opinions and the driving of social divisions (Olan *et al.*, 2024).

According to the results of our 2011 Ox IS study, the majority of UK internet users access the web from the comfort of their own homes. The percentage of internet users has increased somewhat from 70% in 2009 to this year. The amount of time individuals spend online has increased as they participate in more regular and habitual online activities, yet the number of persons utilizing the Internet has not expanded considerably. People's everyday life are increasingly reliant on the Internet, whether for business or leisure. Surprisingly, the percentage of people reading online news has not increased with the increasing number of Internet users from 2009 to 2011, which is a reflection of actual Internet use (Figure 2). We enquired as to whether or not OxIS respondents read an online newspaper. As seen in Figure 1, the percentage increased from 30 percent in 2007 to 57 percent in 2009. In 2011, only 55% of the population said that read a newspaper online, which indicates that this tendency has decreased or perhaps turned around since 2009. Figure 1.2 shows that, in general, males are more likely than women, students and working individuals are more likely to read the news online than retirees and jobless people(Newman, Dutton and Blank, 2014).

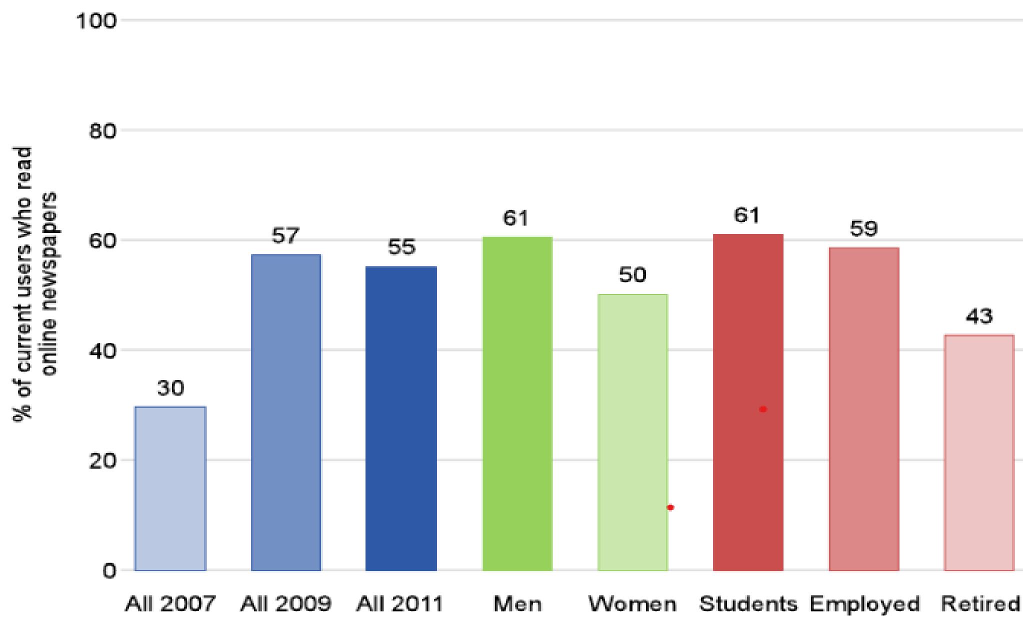

Figure 1.2: Percent of Internet users reading news online by gender and life stage.

1.4 A Typology of Fake News

News Satire

Satire, which refers to mock news shows that usually employ humour or exaggeration to offer viewers with news updates, is the most prevalent operationalization of fake news in the evaluated articles. The Daily Show on Comedy Central in the US is one instance of such a show. (Baym, 2005) But there's a major distinction: instead of calling themselves journalists or newsreaders, the show's hosts refer to themselves as comedians or entertainers. It also promotes themselves as focusing more on giving entertainment than information. The humorous intent of the shows is rather obvious.

Despite its frequent exclusion from mainstream media, satirical programs are proving to be an essential part of the media landscape, according to many studies. Their comedy isn't just for show; it frequently alludes to serious political, social, or economic issues. Essentially, it is a balanced combination of informative and entertaining. discovered that persons who see humorous shows possess a level of awareness of current affairs comparable to those who engage with other types of news media. Satirical shows are recognized for their substantial influence on public debate, attitudes, and political trust (Brewer, Young and Morreale, 2013).

News Parody

The dual use of parody as a tool for both creating and countering fake news highlights its significance. Unlike fake news, parody succeeds when both the writer and the reader or viewer get the joke. A second media that has been characterized by research as "fake news" is parody. It is characterized by a number of similarities to satire, as both rely on humour to engage an audience. The presenting style is also quite similar to that of traditional news outlets. Using fictional characters or events to make a point is what sets parodies and satires apart. Parody, on the other hand, employs wholly fabricated news stories to emphasize the absurdity of issues, rather than offering direct commentary on current affairs through humour (Sinclair, 2020).

News Fabrication

"Fabrication" is the third way that the articles we looked at operationalized fake news. These are publications that don't really include any facts but are released in a way that makes them seem like news items. Unlike parody, there is no presumption on either side that what is being presented is not true. The goal is usually the polar opposite, in fact. In many cases, the manufacturer's goal is to mislead consumers. Anything you make up can end up on the web, in a blog, or shared on social media. When politically biassed institutions publish these articles, seeming to be neutral and giving fair reporting, it becomes more difficult to discern between fake news and real news. To the author's mind, a well-crafted fake news story is one that uses existing memes or biases, just like a parody. These elements are skilfully woven into a story that the audience readily accepts as true, despite the narrative's obvious political leaning. The reader has it even worse when trying to verify claims made by non-news organizations or people that use news presentation styles and techniques to promote false news. When shared on social media, the fact that the material is coming from a trusted source lends credibility. Fake goods can only sell well in situations where social friction already exists. Trust in an organization or person makes people less likely to accept unfavourable news about that institution or person. Approximately 30 million people posted false news stories about Donald Trump on Facebook, whereas 8 million people shared false news stories about Hillary Clinton. Just over half of the people who recalled these tales also bought into their veracity (Allcott and Gentzkow, 2017)

Photo Manipulation

A term "fake news" has expanded to include the practice of fabricating news stories by altering real-life media. Unlike the other categories, which mostly dealt with text-based topics, this one focusses on visual news. The proliferation of digital pictures, robust image editing software, and familiarity with methods has led to an explosion in the frequency of image modification. The effects could be straightforward or complex. Enhancing colour saturation and eliminating extraneous features are two examples of easy modifications (Zubiaga and Ji, 2013).

With an estimated 3.8 billion monthly users, social media platforms like Facebook, Instagram, Twitter, and Weibo facilitate a rapid dissemination of news and public opinion. The exchange of information has been substantially enhanced by these networks. These sites have utilized false content to propagate harmful information and influence public opinion.

Advertising and Public Relations

Press releases passed off as news and advertisements posing as news stories are two other examples of the kinds of content we've seen the label "fake news" used to. "Native advertising" often exaggerates the benefits of the advertised product or person and relies on partial or unreliable information. Nonetheless, it uses the new style to promote its biassed assertions as more legitimate. Another trend is the increase in the use of "clickbait" headlines, which are intended to get readers to "click" and then navigate to a commercial website. As an example, in March 2017, a Facebook ad that went viral featured what seemed to be a news story about a rich Middle Eastern guy who had been pulled down for speeding in the UK, complete with a headline and a photo. He allegedly informed the authorities that his vehicle was worth more than the officer's yearly pay, according to the headline. The article generated unpleasant, even racist, responses, with some people advocating for the guy to be deported. Nevertheless, the user was directed to a marketing website rather than a news article upon clicking on the post. This kind of content often goes under the name "fake news" since it uses the news value to attract attention, but it ends up misleading many people and even making them angry over something that didn't happen (Chen, Conroy and Rubin, 2015).

Table 1.1 displays the outcomes of combining the two scales, which categorize FN definitions from the literature into four distinct groups according to the degree of intentionality and accuracy of the claims made. (Tandoc, Lim and Ling, 2018)

Table 1.1: A typology of fake news definitions.

Level of facticity	**Author's immediate intention to deceive**	
	High	**Low**

High	Native advertising Propaganda	News satire
Low	Manipulation Fabrication	News parody

Source: (Tandoc, Lim and Ling, 2018)

1.5 The Psychology Behind the Spread of Fake News

A psychological issue ingrained in human behaviour coexists with the IT problem of false news spreading via social media networks. Media psychology theories explain how misinformation spreads faster and wider than factual content according to multiple psychological principles.

1. **Cognitive Load Theory (CLT):**

CLT states that people maintain a restricted capability to process information. Users' minds are overwhelmed with the constant exposure to material on social media due to the rapid pace of the platform. Social media users typically use cognitive shortcut tools as well as simple heuristics to judge news content thus making them less likely to evaluate each post properly for genuineness(Paas *et al.*, 2003).

2. **Dual-Process Theory (System 1 and System 2 Thinking):**

According to this theory, human cognition operates through two systems:

- *System 1* is fast, intuitive, and emotional.
- *System 2* is slow, deliberate, and logical.

Social media interactions often engage *System 1* due to time constraints and emotional design (likes, shares, trending topics). This favors the rapid spread of fake news, which typically appeals to emotion and confirmation bias, over factual content that requires deeper analytical thought (*System 2*)(Wixted, 2007).

3. **Social Identity Theory:**

According to this theory people accept and spread information that fits with the core beliefs of their social networks. Fake misinformation which validates group beliefs or political viewpoints receives less attention to critical assessment because it spreads more extensively(Stets and Burke, 2000).

4. **Media Richness and Modality Effects:**

Multimodal content (text, images, videos) increases engagement and memorability. Fake news often leverages rich media to enhance credibility and emotional appeal, making it more persuasive and more likely to be shared—especially when paired with misleading visuals.

1.6 Significance of Fake News Detection

False reports that are created to sway public opinion or disparage an individual are the most basic definition of FN. On social media, false news items frequently receive more views than their real-world counterparts. Consider the social media platform "Facebook," where the top 20 false news items garnered more attention from users than the top 20 real news stories. This finding lends credence to the assertion. The use of social media features such as sharing, commenting, and tagging friends in postings has helped disseminate this news far and wide (Pandey *et al.*, 2022).

With the increase in social media development, a parallel increase in fake news is seen. This news is distracting, obtrusive as well as annoying to the readers. Extensive dissemination of fake news may lead to dangerous impact on society as well as individuals in many ways; (i) It can destroy the authenticity equilibrium of the news environment, (ii) It intentionally convinces readers to acknowledge biased or false information and (iii) The way individuals understand and respond to real news is affected. (iv) This news will dominate the decisions, interests, and opinions of the public.

(v) It will influence the way how people will interact with the real news. (vi) It will destroy the beliefs and faith of people on their experts, authorities, and the government. Vosoughi et al. (2018) Additionally, it will be easier to grasp the significance of fake news identification if you are familiar with the following traits of FN:

- **Fake news volume:**

As there are no verification procedures, it is easy to write FN on the Internet. You can find many web pages whose main purpose is to publish fake stories and news. These websites resemble legitimate news websites and are created to spread false propaganda, misleading information and hoaxes. This is mainly done for political and financial gain. All this happens without the awareness of the website users.

- **Fake news variety:**

Some definitions of fake news include politicians' false statements, conspiracy theories, fake advertisements, misinformation, fake reviews, satirical news, and rumours. A variety of information affects a variety of people and covers every aspect of people's lives.

- **Fake news velocity:**

Most fake news is short-lived. For example, as discussed above, FN propagated during the 2016 U.S. elections no longer exists today as it was removed after the campaign.

1.7 Fake News on Traditional News Media

An issue of FN has been around for a while. Fake news has spread from print and radio to internet news and social media, among other more conventional channels of mass communication. We used to call the problem of FN "conventional FN" before the enormous influence of social media on its production and dissemination. Then, drawing from a variety of theoretical frameworks in the fields of psychology and social science, we will examine how false news impacts both people and social information ecosystems.

Psychological Foundations of Fake News.

People are fundamentally bad at discerning real news from fake. Various theories in cognitive science and psychology provide possible explanations for these occurrences and the impact of harmful FN. Traditional forms of FN mostly target consumers in an effort to take advantage of their particular weaknesses. There are two primary reasons why consumers are naturally vulnerable to FN:

- **(i)** **Naive Realism:** Many customers have the belief that their own opinions are the only valid ones, dismissing others who hold a different view as biassed, ignorant, or illogical (Ross and Ward, 1996);
- **(ii)** **Confirmation Bias:** customers are more receptive to data that backs with their preexisting opinions. This cognitive bias is inherent in the human condition, and it leads consumers to commonly confuse fake news with actual news. On top of that, it's really difficult to fix the misunderstanding once it has taken root. Research in psychology has shown that correcting misleading information (such as fake news) by presenting accurate facts does little to aid in lowering misunderstandings and can even lead to an increase in misunderstandings, particularly among ideological groups (Nickerson, 1998).

Social Foundations of the Fake News Ecosystem.

Several social factors contribute to the spread of FN, which should be considered in the context of the entire news consumption ecology. According to prospect theory, individuals weigh the potential benefits and drawbacks of several options in relation to their current situation before making a final decision. This maximization bias also extends to social benefits, such as maintaining favour with others in the user's close social circle. According to social identity theory and normative influence theory, people's sense of self-worth and identity is greatly impacted by their desire to be accepted and validated by others. As a result, people tend to gravitate towards socially safe options when it comes to sharing and consuming news, even if it's simply fake news (Tversky and Kahneman, 1992).

To apply economic game theory to the rational theory of interactions including fake news, it is possible to utilize a two-player strategy game to model the news production and consumption cycle (Guille *et al.*, 2013). In an effort to shed light on the phenomenon of false news, it is presumed that the information ecosystem is composed of two primary players: publishers and consumers. By incorporating a distortion bias b into the mapping from the original signal s to the final news report a, we can describe the news publishing process as a whole. Here, b = [101] denotes the [left right] biases that impact the news publishing process. This seems to be catching how biassed or twisted a news piece may be to create fake news. Two viewpoints contribute to the publisher's utility:

(i) **Short-term utility**: profit maximization, which is highly connected to the quantity of customers gained;

(ii) **Long-term utility:** their credibility as news sources.

Utility of consumers consists of two parts:

(i) **Information utility:** acquiring accurate and objective information (typically requires more investment expense);

(ii) **Psychology utility**: experiencing confirmation bias and prospect theory as a result of receiving knowledge that backs up their existing beliefs and societal demands. Optimal performance in this game of news consumption is a goal for both the news producer and the reader. An example of fake news would be a situation where the balance is maintained and the short-term utility of publishers exceeds their whole utility, but the psychological utility of consumers exceeds their overall utility. This clarifies the social mechanisms that promote a climate favourable to a dissemination of FN (Shu *et al.*, 2017).

1.8 Fake News on Social Media

The internet has grown ingrained in our daily lives these days. Newsgathering and consumption habits have shifted away from relying on more conventional media like newspapers and television. Undoubtedly, this shift would not have been possible without a proliferation of social media (Bondielli and Marcelloni, 2019).

Social media's affordability, accessibility, and ease of information sharing have drawn users from all over the world. But as a result, fake news spread. As misinformation continues to proliferate on the internet, particularly in places like social media, news blogs, and online newspapers, the detection of false news has recently attracted more interest from the general population and scholarly communities (Allcott and Gentzkow, 2017).

The term "FN" is often utilised to describe news stories that are deliberately and obviously false. Additionally, it uses information that is presented as news but contains factual inaccuracies in order to trick readers into thinking it is real (Bermes, 2021). The post-internet age has rekindled interest among researchers in detecting fake news. Fake news is now one of the worst risks to the country, democracy, and media (Zhou *et al.*, 2019). The political, social, and economic spheres are all severely impacted. There is no shortage of formats for disseminating misinformation and FN. The role that knowledge plays in creating our worldview and developing fact-based critical judgements means that it has a significant influence. The information acquired is used to generate opinions on groups or events. Making educated judgements, however, is impossible when the information being gathered is false, fake, skewed, or contrived. Research shows that nearly all Americans (93%) rely on some type of internet material or technology for their information needs (Nagi, 2018).

The use of social media considerably facilitates the propagation of inaccurate and deceptive information. In the 2016U.S. presidential election, for example, the Republican Party ran a smear campaign against Hillary Clinton. Unfortunately for Hillary Clinton, this caused the American people to think she was wrongfully charged, which ultimately cost her the election (Waszak, Kasprzycka-Waszak and Kubanek, 2018). In 2017, there was another example when incorrect information about the suspect was disseminated on social media with news of the Lavages massacre, which killed at least 59 people and injured more than 500 (Meese, Frith and Wilken, 2020). Misinformation in this field may have a major influence on people's lives as more and more individuals are seeking health-related news online. Therefore, this is considered one 2 of the most significant challenges of today. In recent years, misinformation regarding health has had a huge influence (Allington *et al.*, 2021).

1.9 Understanding the Creators Behind Fake News

- **Non-humans:**

Social bots are computers programmed to mimic human behaviour on social media platforms, including content creation and user interaction. Some of the bots may spread real news but most of them spread misinformation, malware, spam, and rumors. For example, the creation of social bots was done during the elections time in the US to support Clinton or Trump, making use of multiple tweets which gave references to many fake news websites. Another non-human fake news spreader is “Cyborgs”. Humanoid robots or bots that help humans do tasks are known as cyborgs. After signing up, these accounts have the ability to tweet many times and join in any social community. Even cyborgs spread misleading information, damaging the trust and belief of social media users. These are some of the prominent non-human creators of fake news.

- **Humans:**

People are one of the main sources of false news. Even if fake news is spread automatically or manually the main distributors and creators of the fake news are the humans. For example, one agent of the FBI who is suspected in the e-mail leak of Hillary is found dead in his apartment. This information is entirely false, but many users have spread this news by forwarding and sharing this information many times. The followers and friends of this user further spreads this news multiple times, these are called next-generation spreaders and this spreading results in an echo chamber propagating a spread of FN. This study clarified a definition of false news, an importance of spotting it, and the different kinds of people that distribute it. Let us discuss detection methodologies for fake news.

1.10 Fake news detection approaches.

Three groups of techniques for spotting fake news are used in a recent categorization. The three types of existing methods—content-based, feedback-based, and intervention-based—further subdivide each category. Literature reviews on the topic of detecting FN in online social networks do, however, reveal that current research may be broadly grouped according to two main features that the majority of writers examine and utilize to establish a suitable solution. The content-based component is concerned with the substance of the news item, while the contextual aspect is concerned with the news post's context. Both of these factors may be seen as significant sources of extracted information utilized for FND(Sharma *et al.*, 2019). Fake news detecting methods are displayed in Figure 1.3.

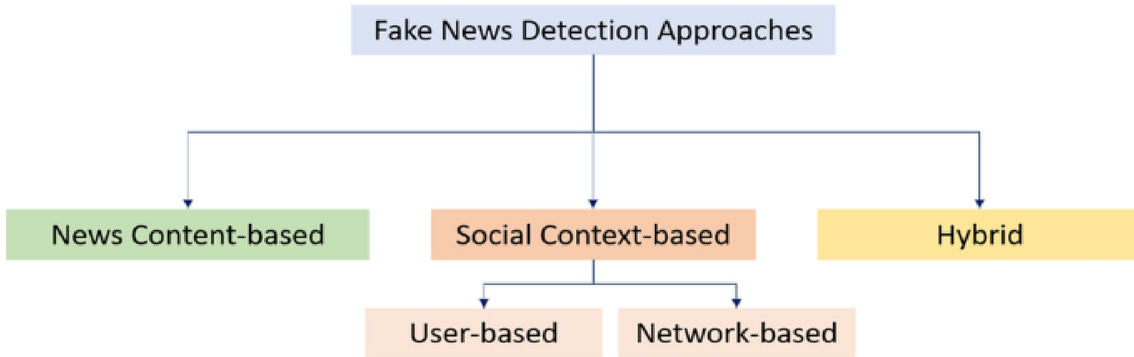

Figure 1.3: Classification of fake news detection approaches

News Content-based Category

A news content-based strategy is one that studies and uses the news content to find answers; this technique uses content information, which is information collected from the news item, to detect fake news. Some minor differences may be seen in the news material, which includes the headline, text, image/video components, and source (Kaur, Kumar and Kumaraguru, 2020).

Researchers employ content-based detection cues, which include both textual and multimedia signals, extracted from the news item. There are two types of news cues: those based on multimedia (pictures and videos) and those based on text (components in the news itself). Vereshchaka et al. (2020) Several well-known methods exist for detecting FN in news content, which encompasses both text and multimedia/images. These methods include ML, DL, fact-checking, crowdsourcing, and blockchain. A quick summary of these approaches is shown in Figure 1.4 (Mahabub, 2020).

The majority of researchers in this area use artificial intelligence (AI) methods (ML, DL, and NLP models) to increase prediction accuracy. Some employ alternative methods including blockchain, crowdsourcing, and fact-checking. Specifically, the approaches based on AI and ML in this field strive to extract traits from news data for utilization in training and content analysis assignments down the road. The many kinds of data deemed pertinent for the study in this instance are the extracted characteristics.

Automatic fake news detection relies on a small amount of data, and feature extraction is a great way to make that data smaller. The goal of this method is to enhance classification performance by selecting a subset of characteristics from the whole set (Yazdi *et al.*, 2020).

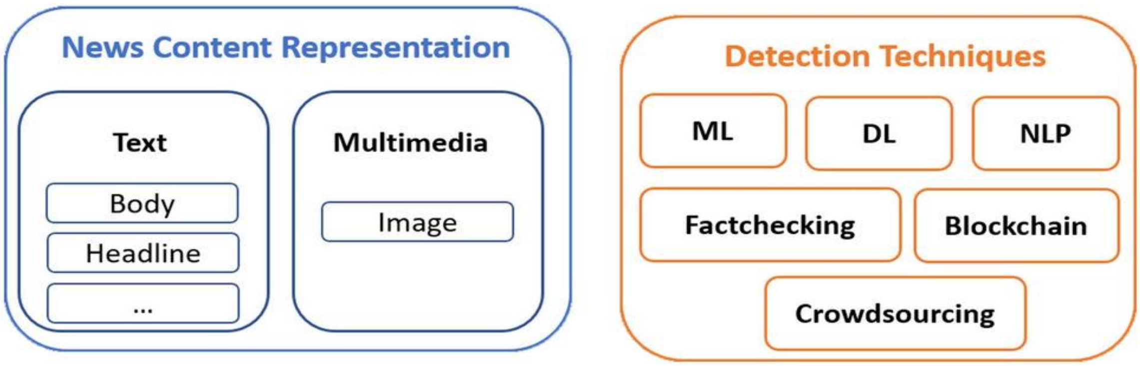

Figure 1.4: News content-based category: news content representation and detection techniques.

Social Context-based Category

In contrast to solutions focused on news content, social context-based approaches try to understand a sceptical social climate around online news. The social context-based approach to identifying FN includes methods that depend on contextual qualities, such as information related to the news post's context. Additional information that may be utilized to spot fake news is provided by these criteria, which are obtained from social context. Data that is not directly related to the false news article, called "contextual data," might be crucial for automated systems to identify fake news. Checking the article's legitimacy and publisher's credentials, finding the publishing date or any supplemental materials, and checking to see whether other online news sites are covering similar or identical topics are all examples of pertinent contextual information (Zhang and Ghorbani, 2020).

An component of social context may be either user-based or network-based. Both types may be employed for training and context analysis in AI and ML-based systems.

The term "user-based aspects" describes data collected from people using OSNs, such as their profiles (Nyow and Chua, 2019).

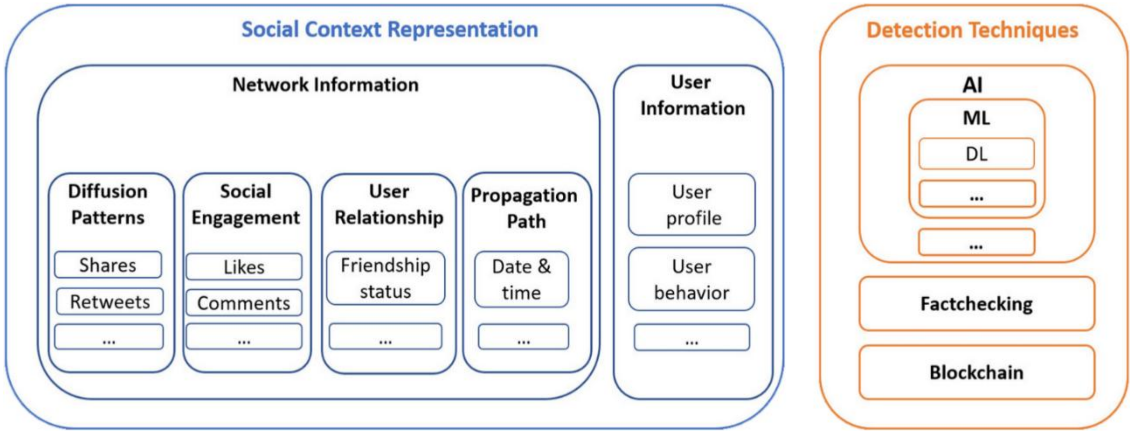

Figure 1.5: Social context-based category: social context representation and detection techniques

Hybrid approaches

The majority of academics are neglecting the benefits of combining content- and context-based methodologies in favour of a singular approach. This is because some individuals hold the view that conventional fusion procedures still have significant restrictions brought forth by preexisting feature correlations and semantic conflicts. Because of this, some academics are proposing methods that rely on content-based data extraction, while others are including data based on social context (Wu and Rao, 2020).

Nevertheless, identifying false news using a single attribute has proven to be difficult. As a result, contemporary trends use a hybrid approach that draws from both social context and news content (Ruchansky, Seo and Liu, 2017).

1.11 Need for fake content detection system

From manipulating financial markets to supporting political 7 extremism, its power extends well beyond the political sphere. Below are some of the repercussions that have been seen throughout the propagation of fake information.

- **Political Influence:**

During the 2016 US presidential election, the spread of fake news was significantly boosted. It came to light that gangs of Macedonian adolescents were making Trump-supporting material with the intention of profiting from its shareability. The Russian propaganda operation was discovered to have been fabricating pro-Trump fake news to make it seem like there was a large grassroots movement genuinely behind Trump's campaign. The impact of these pursuits is difficult to measure. In any case, it's clear that the rise of FN as a tool to influence politics is a grave concern (Canini, Suh and Pirolli, 2011).

- **Encourages Mistrust in Legitimate Media Outlets:**

Politicians have used the public's heightened knowledge of fake news to their advantage by smearing news organizations that report critically on their policies. Politicians in the United Kingdom and the US have used the phrase "fake news" to cast doubt on press coverage of their opponents' or their own policies. Also, the worth of expert knowledge has been eroded by the proliferation of fake news (Caetano *et al.*, 2018).

- **Influences the Financial Markets:**

Fake news of French-English peace in 2015 caused share prices on the London Stock Exchange to jump 5% almost instantly. In order to boost trust in specific cryptocurrencies and so inflate their value for the advantage of currency holders, the majority of this news was disseminated via platforms such as Twitter and Telegram. (Gupta, Lamba and Kumaraguru, 2013)

- **Bad for Business:**

The prevalence of politically driven fake news operations aimed at large businesses is on the rise.

- **Damaging to Personal Reputation:**

Everyone is acquainted with the celebrity death hoax as an example of a spread of FN, but there are many more forms of false articles produced about celebrities. A number of famous people were the targets of discrediting fake news operations. For instance, Amitabh Bachan was pronounced dead during his abdominal surgery.

- **Defaming of celebrities:**

Deep fake technology developed by amateurs has advanced to produce virtual actors for films. The distinction between actual and virtual movies is invisible to the human eye. Because their fake dubbed audio and video snippets make unexpected or irrelevant claims, this causes problems for well-known celebrities. Hence in charge of responding to journalists, the media, etc.

- **Damage to Society:**

In 2017, an old Chinese woman was unjustly accused of being part of a squad that kidnapped children. Following verification and inquiry, a woman was recognized as a frequent client at a retail centre. The woman was later proved innocent, but it was of little use because the harm had already been done to her. The old woman was frightened to leave her house and avoided social situations. The next section discusses the numerous uses of fake content detection systems. Fake content is posted on a variety of social media sites (Wang, 2010).

1.12 NLP techniques used in fake news detection

Computers can now comprehend, analyze, alter, and maybe even generate human languages due to NLP, a state-of-the-art subfield of ML. This method has involved a number of activities, including feature extraction techniques, word embedding, and pre-processing. Many approaches for identifying fake news begin with data pre-processing. It's utilized for describing cryptic qualities, managing missing words, binarization attributes, and generating intricate structures containing attributes. A number of visualization procedures are helpful during data pre-processing. Data pre-processing eliminates noise and saves storage and processing time. Second, the process of word vectorizing is converting text or words into a set of vectors. In addition, TF-IDF and a word bag are often used by several ML frameworks for a purpose of detecting fake news. Recently, methods for spotting fake news have relied on pre-trained word-embedding models, such as Glove and word2vec, which can learn larger datasets (Nirav Shah and Ganatra, 2022).

1.13 Consequences of Fake News

An emergence of human civilization has been marked by the frequent appearance of fake news. The global media ecosystem and contemporary technology, however, may be used to spread fake news. The social, political, and economic spheres are all impacted by fake news. Fake news and information, however, come in various forms. Fake news has a big impact on people's perceptions all over the globe, even though it may lead to bad decision-making and the misuse of important decisions. Similar to how you can't rely on this inaccurate, twisted, or fake information found online to make sound decisions. The health of innocent people, the integrity of democracies, and the economy are impacted most severely by fake news (Aslam *et al.*, 2021).

- ***Democratic Impact*:**

A significant impact of FN on the election has prompted media discussions about it. Many also see it as a fundamental issue for democracies. So, finding and stopping the spread of FN is really essential.

- ***Financial Impact:***

False news has become an increasingly complicated issue that affects many parts of society. To boost their own earnings, dishonest businesspeople could spread fake reviews or news. The reputation of a company can be severely damaged by fake information.

- ***Impact on Health:***

Internet users are more likely to look for news on health. There has been an alarming rise in the prevalence of false news in the healthcare sector in recent years, putting many people's lives at risk. Social media settings have therefore led to several regulatory reforms that impact doctors, politicians, and health advocates by restricting or outlawing the dissemination of false information in articles pertaining to health.

- ***Impact on Innocent People:***

Rumours have the power to influence certain men. People harass these people on social media. Persons may also face threats and insults with actual consequences (Nirav Shah and Ganatra, 2022).

1.14 Applications of Fake Content Detection System

Analysis for detecting FN has several potential uses. A few of the more crucial uses are outlined below.

i) Control of fake spread during elections:

During the general elections in India in 2019, fake news was widely disseminated. The campaign for the election was marred by the dissemination of false information. Many people in India utilized WhatsApp as a propaganda tool to spread false information, which led to the first WhatsApp elections. After examining the accounts using technology that identify fake news, the unauthorized material on Twitter and Facebook was deleted. (Garimella and Tyson, 2018)

ii) Corona pandemic:

False information about COVID-19 may be found in social media postings about home remedies that have not been confirmed to work, fake warnings, and conspiracy theories. Fake news regarding the coronavirus epidemic led to the arrest of two persons. The prime minister of India issued a strong warning about a spread of FN on the COVID-19 epidemic. Debunking such false material is a collaborative effort between many scientists who are employing human assistance and creating and deploying fake instruments (Zhang *et al.*, 2016).

iii) Terrorism:

The widespread user base of social media sites like Facebook, YouTube, and Twitter has made them a prime tool for terrorist organizations and individuals looking to disseminate their messages. There have been efforts by some countries and organizations to block terrorist groups from using social media. The low cost, ease of usage, and large audience reach of social media make it an attractive platform for terrorist organizations. After observing their social media networks for a while, professionals can examine their aggressive, hypocritical ideals through crowdsourcing.

iv) Natural calamities:

People on social media have a tendency to trust posts about natural catastrophes or crises and repost them in the hopes of reaching a large number of other users. Sadly, some malicious people are cognizant of this tendency and intentionally promote it by uploading bogus information, such as spam and fake messages. Social media is regularly used to spread spam and fake photographs after natural catastrophes (Batchelor, 2017).

v) Fighting riots:

Malicious members of online social networks like to utilize fake identities to transmit spam, perpetrate fraud, and engage in other forms of system abuse. In order to extend their operation and reach as many authentic members as possible, a single malevolent actor may generate dozens to thousands of fake identities. Protecting genuine members and keeping the network trustworthy requires immediate detection and action on rogue accounts. Since communal conflicts may cause harm to society in the form of personal loss, faith, the economy, and national pride, the government and police are continually on the lookout for them. Therefore, malevolent members of online social networks prefer to utilize fake identities in order to transmit spam, conduct fraud, or otherwise misuse the system.

1.15 Open challenges to detect fake content

Finding fake news items on social media sites that appeared as multimodal signals presented a variety of difficulties. The next section discusses a few of these difficulties.

i) Traditional algorithms:

In many sectors, a number of methods have been created for detection. Unfortunately, there aren't any comprehensive publicly available databases that include false information, so it's hard to compare hoaxes, social media gossip, and fake reviews. This stops benchmark comparisons across various algorithm categories.

ii) Unstructured data:

Social media networks see massive amounts of data, with unstructured data (pictures with text, audio, and video) accounting for more than 70% of that data. Newspapers, journals, verified and unverified news websites, etc. are some of the sources that these platforms compile information from. The analysis becomes more complicated when dealing with such unstructured data types as internet information, which can be viewed in many formats.

iii) Unavailable datasets:

There is a lack of deep fake benchmark datasets pertaining to well-known politicians, necessitating a substantial effort investment in cataloguing the actual video speech segments. There was never any intention of implementing a conventional splitting technique prior to adopting a broad approach to partition movie collections at the pixel or frame level. When collecting source and destination video clips, several limitations need to be considered in order to construct deep fake face-swap clips.

iv) Mode of education:

A significant untapped research topic is how to mitigate the effects of false information. Recent studies have shown that human detection skills are enhanced when the public is made aware of possible manipulation techniques used in false information. In order to discover ways to educate individuals not to believe fake information and to scale these tactics to millions of social media users, further study is needed.

v) Third-generation deep fake dataset:

A third-generation deep fake dataset presents significant challenges when trying to implement the models. Not only are third-generation datasets more abundant (by a factor of more than ten) than second-generation datasets, but there are also of higher quality (containing the consent of persons featured in the dataset).

vi) Match with false content:

A knowledge base containing accurate information may be used to verify false information. There are a lot of challenges with this technique, including making sure the data is good and efficiently creating and maintaining this database of knowledge.

vii) Appropriate feature extraction:

The development of methods for autonomously extracting information from natural language and other sources is essential. The last step is to ensure that the extracted data is consistent with what we already have in our knowledge base by using information-matching techniques (Rao, Verma and Bhatia, 2021)

1.16 Introduction to deep learning

DL is a subfield of ML. In an effort to make advantage of data abstraction at a high level, this technique employs several nonlinear transformations or processing layers with complicated topologies. One kind of ML technique is DL, which uses data characterization to make predictions. (Schmidhuber, 2015) A subfield of ML called DL learns high-level data abstractions using hierarchical structures. This new method is quickly gaining traction in established AI fields including Computer Vision, NLP, semantic parsing, transfer learning, and many more. The tremendous advancements in machine learning algorithms, the drastically reduced cost of computer gear, and the greatly expanded processing capacities of chips (e.g., GPU units) are the three primary causes behind the current boom of deep learning (Ren and Xu, 2015).

A great deal of writing evaluating and analyzing various deep-learning techniques has appeared in the last few years. Among these, a chronological outline of significant influences and technical advancements was used. At the same time, we proposed several new research directions and examined the challenges in DL. The ability of deep networks to simultaneously execute discrimination and feature extraction makes them ideal for computer vision applications (Bengio, 2013).

Deep learning is a notion that is related to shallow learning. In the 1990s, shallow ML models like LR and SVM were established. These shallow ML models, as seen in Figure 1.6, feature either no hidden layer nodes or only one layer. The foundation of deep learning is a network of hidden layers. NN with many layers is the backbone of deep learning. To learn extremely abstract data attributes, deep learning simply passes the input from one layer to the next.

There are two main types of Neural Network models: A bottom-up approach is used by the discriminative model, which passes data via the input, hidden, and output layers. Supervised training uses them for tasks like regression and classification. A generative model, in contrast, is top-down and data flows counter-clockwise. These are useful in situations involving probabilistic distributions and unsupervised pre-training.

When given an input x and its associated label y, a discriminative model will learn a distribution of probabilities p(y|x), which is the probability of y given x. On the other hand, a generative model will learn the joint probability of p(x,y), which allows one to forecast P(y|x) from the input data (Ng and Jordan, 2002). There are two main techniques to data analysis: discriminative methods, which are used when labelled data is available, and generative methods, which are used when labelled data is not (Bishop and Lasserre, 2007). There are essentially three distinct kinds of training:

Unsupervised learning does not make use of feedback as it does not have a labelled data set, in contrast to supervised learning which uses labelled data to train the network. Standard supervised learning techniques may be used to fine-tune neural networks that have been pre-trained in unsupervised learning using generating models like RBMs. The test data set is then subjected to it in order to identify any patterns or classifications. Big data has contributed to deep learning's advancements by providing a wealth of diverse data. There is disagreement about whether supervised learning is superior to unsupervised learning, which defies our natural tendency. Each has advantages and applications. T. Zhou et al. (2017) provided evidence that unsupervised learning outperforms supervised learning on unstructured video sequences for predicting monocular-depth and camera motion. In order to enhance performance, Deep Belief Networks (DBM) and other Modified Neural Networks employ supervised learning using labelled data and unlabelled data, respectively, as explained by Xue-Wen Chen and Xiaotong Lin (2014) The development of an automated feature extraction system from high-dimensional data spaces, both labelled and unlabelled, is a tough task. Yann LeCun et al. proposes that a potential answer to this issue might be the integration of supervised and unsupervised learning (LeCun, Kavukcuoglu and Farabet, 2010). To enhance unsupervised learning, semi-supervised learning incorporates both supervised and unsupervised data, creating a hybrid technique. Overfitting and early convergence are two big obstacles that DNN and training algorithms must overcome. When a DNN's bias and weights settle into a locally optimum state without taking into account the global minima of the whole multidimensional space, this is called premature convergence.

Overfitting refers to a condition in which deep neural networks become too specialized to a specific training dataset, rendering them inflexible and less suitable for any alternative test dataset.

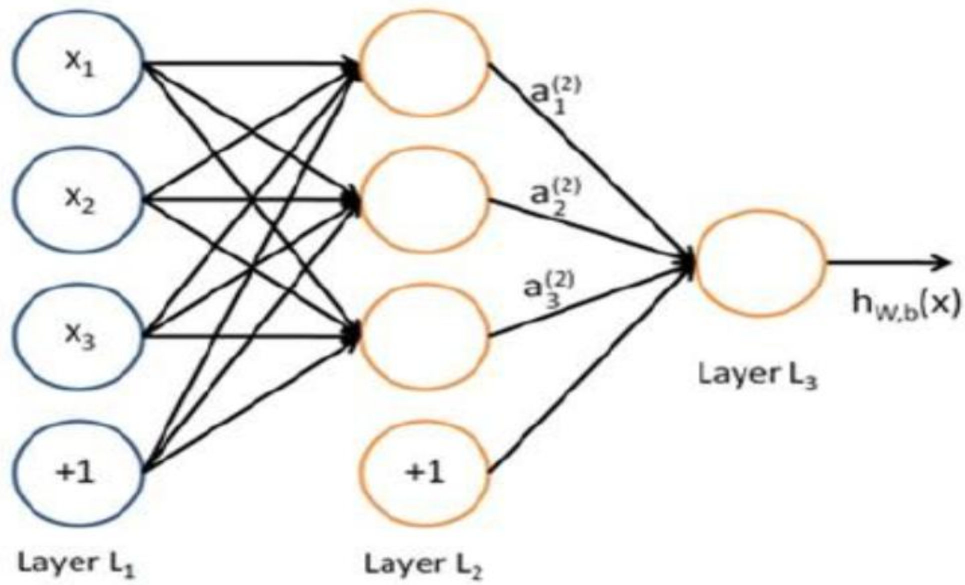

Figure 1.6: A single-layer neural network

Similar to ML, DL may be split into 3 categories: unsupervised, semi-supervised, and supervised. CNN, GAN, Deep Belief Networks, and Restricted Boltzmann Machines are now included in the standard DL paradigm.

1.17 Classification of Neural Network

There are several ways to categorize neural networks.

1. Feedforward Neural Network (FFNN)
2. Recurrent Neural Network (RNN)

3. Radial Basis Function Neural Network
4. Kohonen Self-Organizing Neural Network
5. Modular Neural Network

A FFNN only allows data to go in one way, by an input layer all the way to an output layer (via any hidden nodes, if any). Neither a circle nor a loopback will be formed by them. An example of a multilayer FFNN that uses values and functions produced along the forward pass path is shown in Figure 1.7. Z is the input weighted total, and y is the layer-specific non-linear activation function f of Z. As the subscript letters indicate neighbouring layers, W denotes the weights among the two units, and b stands for the unit's bias value.

In contrast to FFNN, RNN processing units form a cycle. Oftentimes, the sole layer in a normal network structure is the one directly below it, and it receives the output by the previous layer. Consequently, a feedback loop is formed when the layer's output becomes its own input. Because of this, the network can remember its previous states and use them to improve its current performance. This difference allows RNNs to analyze time-phased input data sequences, unlike feedforward neural networks, and create sequences of output values; this is useful for tasks like as voice recognition and frame-by-frame video categorization. This property makes RNNs ideal for these types of applications.

Figure 1.7 shows the progression of an RNN's unrolling over time. As an illustration, consider an input consisting of a series of three-word phrases. Here, we have a three-layer RNN where each word stands for a layer that has been unfolded or unrolled three times.

The diagram's mathematical interpretation is as follows: An input at time t is denoted by xt. The parameters that are learnt and used by all phases are $U, V,$ and W.

Ot represents the result at time t. A state at time t is denoted by St, which may be calculated as follows, where f is an activation function, e.g., ReLU.

$$St = f\ (Uxt + Wst{-}1) \ldots\ldots\ldots \quad (1.1)$$

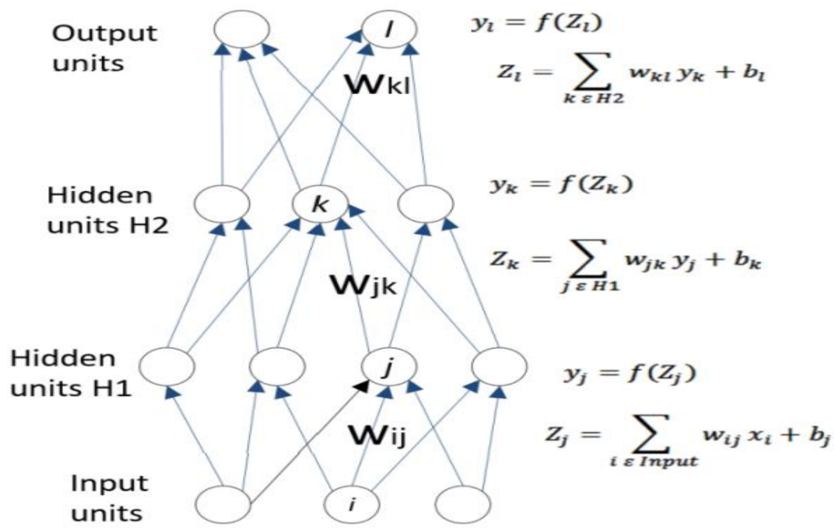

Figure 1.7: Feedforward neural network.

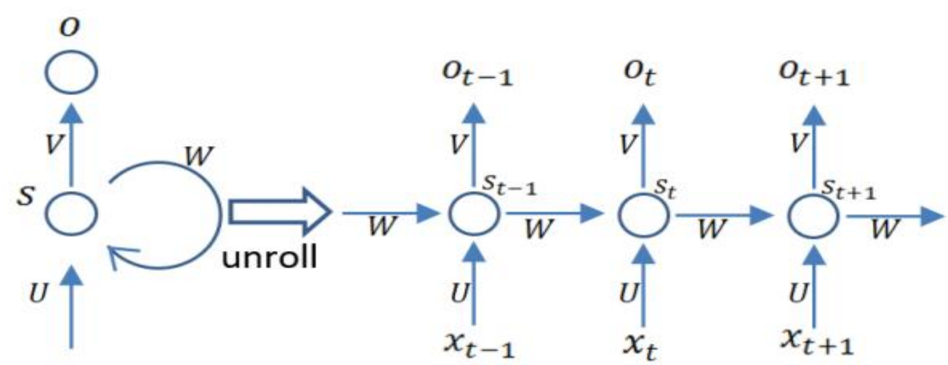

Figure 1.8: The unrolling of RNN in time.

Among the many uses for radial-basis-function neural networks are categorization, function approximation, and timeseries prediction. It has three layers: input, hidden, and output. A radial basis function, with each node standing in for a cluster centre, is included in the hidden layer, which is itself constructed as a Gaussian function. As the network learns to assign inputs to specific nodes, inference and classification are carried out by the output layer, which combines radial basis function and weight parameter outputs (Buhmann, 2000).

The model of a Kohonen self-organizing neural network is trained using input data in an unsupervised manner. Completing its architecture are two interconnected layers: input and output. A two-dimensional grid is used to organize an output layer. In an absence of an activation function, a weight stands in for a property (the position) of a node in an output layer. An output layer node are all given their weighted Euclidean distance by an input data. The closest node and its neighbours' weights are updated using the following formula to make them more comparable to the input data (Akinduko, Mirkes and Gorban, 2016).

$$wi\ (t + 1) = wi\ (t) + (t)\eta j{*}i(x(t) - wi\ (t)) \ \ldots\ldots\ldots\ldots (1.2)$$

Where (t) is the data that was entered at time t, $wi\ (t)$ represents the ith weight as of time t and $\eta j{*}i$ is a function that connects an $ith\ and\ jth$ nodes in the neighbourhood.

A modular neural network divides a big network into more manageable, standalone neural network units. The individual tasks carried out by the smaller networks are then integrated into the network's overall output (Chen, 2015).

The following are some common approaches to implementing DNNs:

1. Sparse Autoen coders
2. Convolution Neural Networks (CNNs or Conv Nets)
3. Restricted Boltzmann Machines (RBMs)
4. Long Short-Term Memory (LSTM)

Autoencoders are a kind of NN that learns encoding or features from a dataset in order to decrease dimensionality. A kind of autoencoders known as a sparse autoencoder has certain units that either produce values that are almost zero or are dormant and donot fire. DCNNs employ a multi-layer architecture that interacts with the input (image pixel values) to extract features. NLP, image recognition, and recommender systems are some of the many places CNN finds use. Using RBM, one may learn the data set's probability distribution.

Backpropagation is the training method used by all of these networks. The weights in backpropagation are adjusted using gradient descent, which reduces errors by taking the partial derivative of the error with respect to every weight.

1.18 DNN Architectures

Numerous nodes organized into several layers make up a DNN. Various designs have been created to address issues in various fields or for specific purposes. Computer vision and image identification make heavy use of CNNs, whereas time series issues and forecasts often use RNNs. However, there is no obvious winner for generic issues like classification since there are a number of variables that might influence the architectural selection. 179 classifiers were investigated, but the best results were found using parallel random forest, or parRF_t, which is just a parallel implementation of a DT variant. Here are three of the most popular deep neural network designs (Fernandez-Delgado *et al.*, 2014).

1. CNN
2. Autoencoder
3. Restricted Boltzmann Machine (RBM)
4. Long Short-Term Memory (LSTM)

Convolution Neural Network

The CNN is the go-to for computer vision tasks like picture identification and video recognition because it simulates the brain's visual processing architecture. NLP, drug development, and other fields also make use of it. The standard design of a CNN, as seen in Figure 1.9, consists of many convolution and sub-sampling layers, a fully connected layer, and a normalising_layer (like a SoftMax function). In Figure 1.9, we can see the famous 7-layer LeNet-5 CNN architecture that was developed for number recognition by (Lecun *et al.*, 1998). A network of convolutional layers extracts features with increasing granularity as it progresses from input to output. Layers that carry out categorization and are fully linked come after the convolution layers. In between each convolution layer, you may see a sub-sampling or pooling layer. CNN accepts as input a 2D $\boldsymbol{n\ x\ n}$ image with pixels.

Filters or kernels are collections of 2D neurones that make up each layer. Each CNN feature extraction layer's neurones are not linked to every neighbouring layer's neurone, in contrast to other neural networks. The input image or feature map of the previous layer partially overlaps with the spatially mapped fixed-sized neurones, therefore only those neurones are connected to the layer. This portion of the input is known as the local receptive field. There is less chance of overfitting and less training time when there are fewer connections. The neurones that make up a filter are all hardwired to the same number of neurones in the feature map or input layer above them and have identical weights and biases. The learning process is accelerated and the network's memory needs are reduced by these components. Consequently, an input images are examined in various regions by each neurone in a particular filter, all in search of the same pattern. (Lecun and Bengio, 1995) Network size is decreased via sub-sampling layers. Beyond that, it successfully lessens the network's vulnerability to picture scale, shift, and distortion when combined with shared weights (inside the same filter) and local receptive fields. The process of sub-sampling is accomplished using methods such as max/mean pooling and local averaging. After each layer of a CNN has completed its connectivity, the last layers are in charge of performing the classifications. In order to construct deep CNNs, it is possible to employ several sets of convolution layers that share weights and perform sub-sampling. While maintaining locality, using fewer parameters, and being invariant to small changes in the input picture, CNNs provide high-quality representations due to their deep nature. (Taylor *et al.*, 2010).

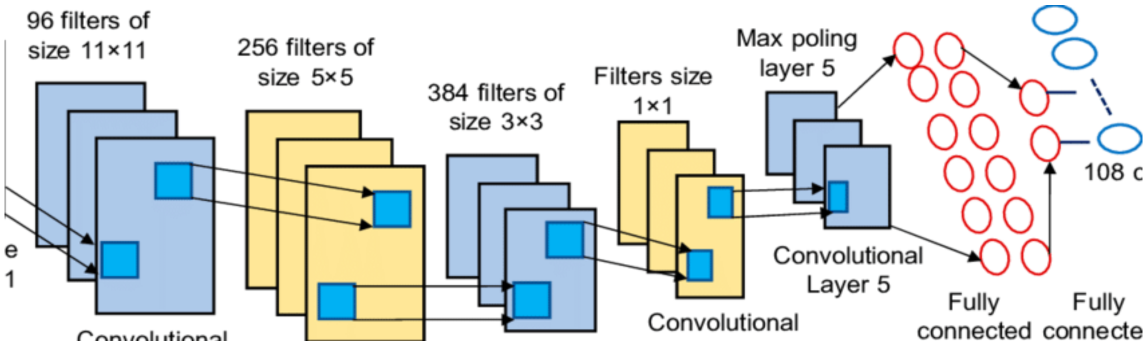

Figure 1.9: 7-layer Architecture of CNN for character recognition

The most popular variants and applications of the CNN architecture are listed below.

1. Alex Net:

When tested against all of the standard methods in ML and computer vision, Alex Net outperformed them all in terms of recognition accuracy. There was a dramatic uptick in interest in deep learning after this seminal work in Computer Vision and ML for image identification and categorization. Figure 1.10 shows the Alex Net architecture. A first convolutional layer uses Local Response Normalization (LRN) to perform convolution and max pooling using 96 separate 11×11 receptive filters.

Max pooling operations in 3×3filters require a stride size of 2. In the second layer, 5x5 filters are used to carry out the identical processes. Each of the three convolutional layers utilizes 384, 384, and 296 feature maps with 3×3 filters, correspondingly. Two dropouts Fully Connected (FC) layers are concluded with an application of a SoftMaxLayer. Two parallel-training networks sharing a same topology and having an identical number of feature mappings are used for this model. This network introduces two new ideas: dropout and Local Respons Normalization (LRN). There are two methods to apply LRN. One option is to apply it to a single channel or feature map. Here, a N×Npatch is chosen by a same map and subsequently normalized using a values in the adjacent neighbourhoods. Secondly, LRN may be used on feature maps or channels, which are essentially the same thing: a neighbourhood along a third dimension, but with only one pixel or position (Krizhevsky, Sutskever and Hinton, 2012).

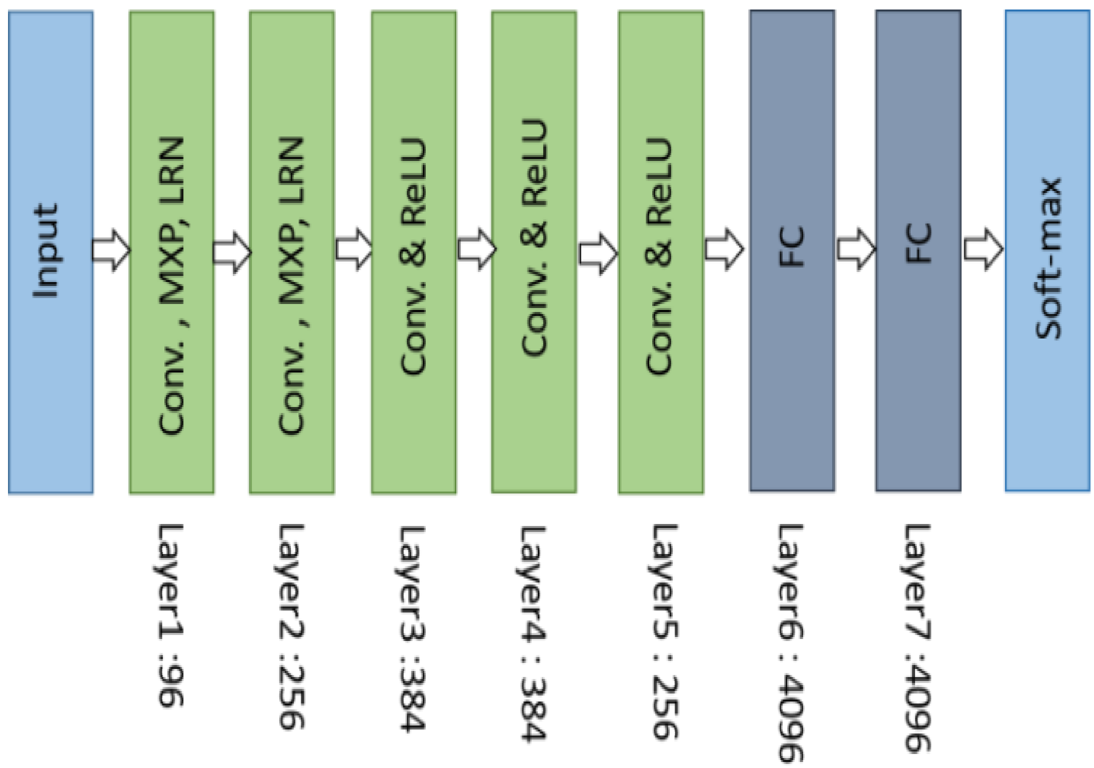

Figure 1.10: Architecture of Alex Net: Convolution, max-pooling, LRN and fully connected (FC) layer.

2. Inception:

It was suggested to use the pre-trained model Inception-V3. One of the leading authorities on hardware in the industry trained this model, which includes over 20 million parameters. Every symmetrical and asymmetrical building block of the model has a different mix of layers, including convolutional, average, max pooling, Concat, dropout, and fully connected ones. Furthermore, this model's activation layer input is often subjected to batch normalization. SoftMax is used for classification. Figure 1.11 displays the Inception-V3 model's schematic diagram (L. Ali *et al.*, 2021).

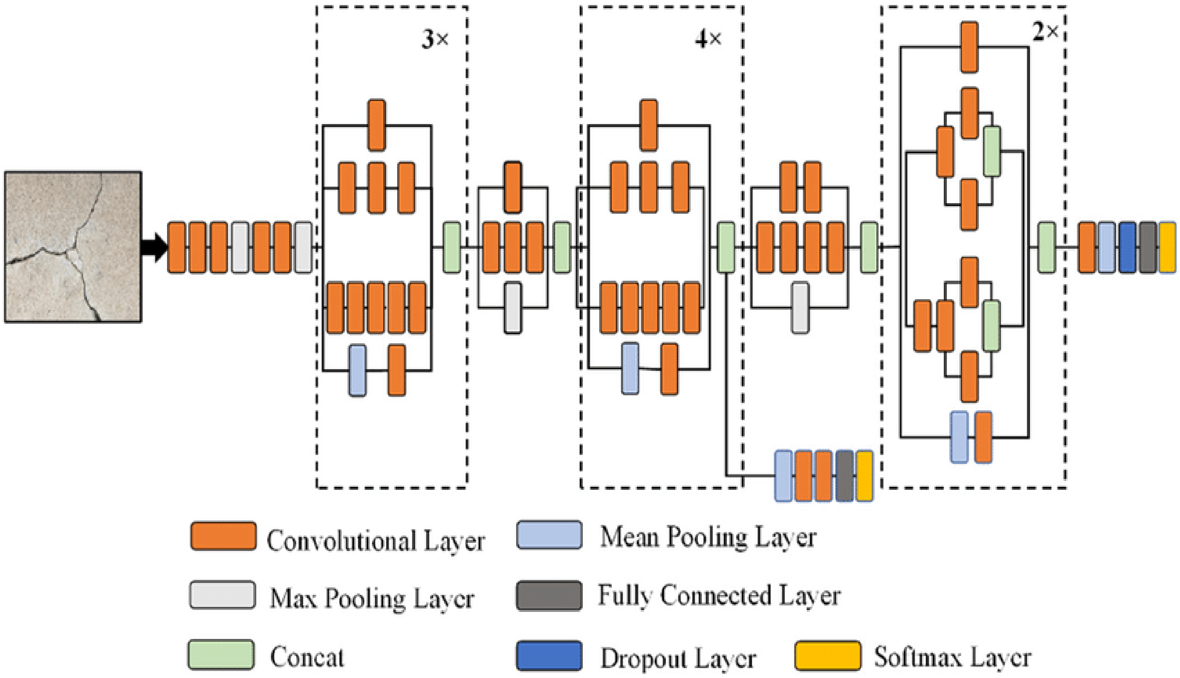

Figure 1.11: The structure of the Inception-V3 model's architecture.

3. Res Net:

The Residual Network design, ResNet, was crowned the victor in the 2015 ILSVRC. Designed to solve the vanishing gradient issue, Kaiming's ResNet allows for the creation of ultra-deep networks. A wide variety of layer counts, from 34 to 50 to 101 to 152 and even 1202, are used while developing ResNet. One completely linked layer capped off the 49 convolution layers that made up the well-known ResNet50. There are a total of 25.5 million weights and 3.9 billion MACs in the whole network. (He *et al.*, 2016). Figure 12 displayed the ResNet architecture.

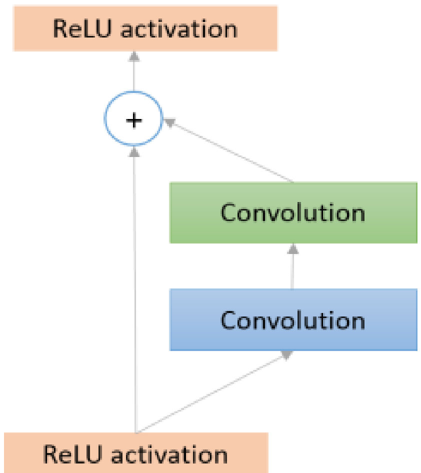

Figure 1.12: Architecture of Res Net.

4. VGG:

In 2014, the ILSVRC was won by the Visual Geometry Group (VGG). The key finding of this study is that deeper networks perform better in CNNs when it comes to identification and classification. The VGG architecture's two convolutional layers use the ReLU activation function. Following the activation function are many completely connected layers that use a ReLU activation function, as well as a max pooling layer. A model concludes with a classification SoftMax layer. A 3x3 convolution filter with a stride of 2 is used in VGG-E. These three VGG-E models were suggested: VGG11, VGG-16, and VGG-19. Eleven, sixteen, and nineteen layers were present in the models, correspondingly (Simonyan and Zisserman, 2014).

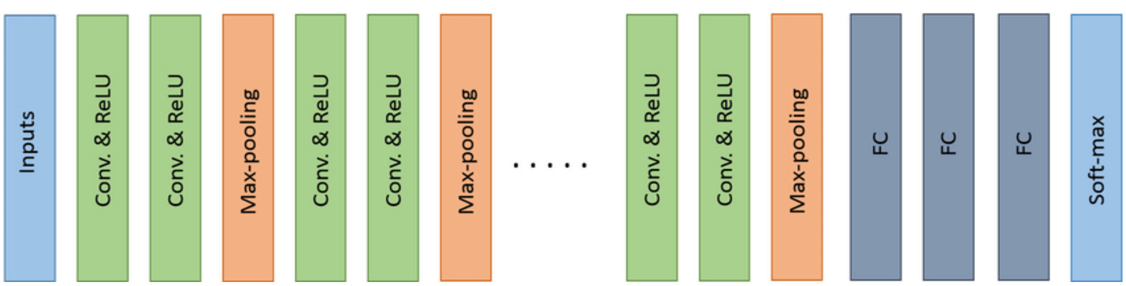

Figure 1.13: The basic building block of VGG network: Convolution (Conv) and FC for fully connected layers.

5. DCGAN:

GAN method was used with CNNs to enhance the standard of synthetic creation, particularly for visual content. One area where CNNs have shown to be very effective is in Computer Vision tasks pertaining to image recognition. The DCGAN method was proposed as a result of these two factors coming together. (Radford, Metz and Chintala, 2016) A series of convolutional procedures including spatial up-sampling processes is used by the DCGAN approach to enhance generation. In an effort to mitigate the mode collapse problem in GANs, the DCGAN design was first proposed. The generator experiences mode collapse if it becomes biassed towards a small number of outputs and stops producing distinct outputs for each dataset variation. Other GAN designs rely on the DCGAN architecture, which has shown to be reliable throughout training for image creation. According to the stable training architectural requirements, the pooling layers of a classic CNN should be replaced by stridden convolutions in the discriminator model and fractionally stridden convolutions in the generator model. To make your data smaller, you may use a deep learning method called a stressed convolution. The data size is increased using fractionally-strided convolutions, a deep learning approach (Radford, Metz and Chintala, 2016). Figure 1.14 DCGAN (Deep Convolutional Generative Adversarial Network)

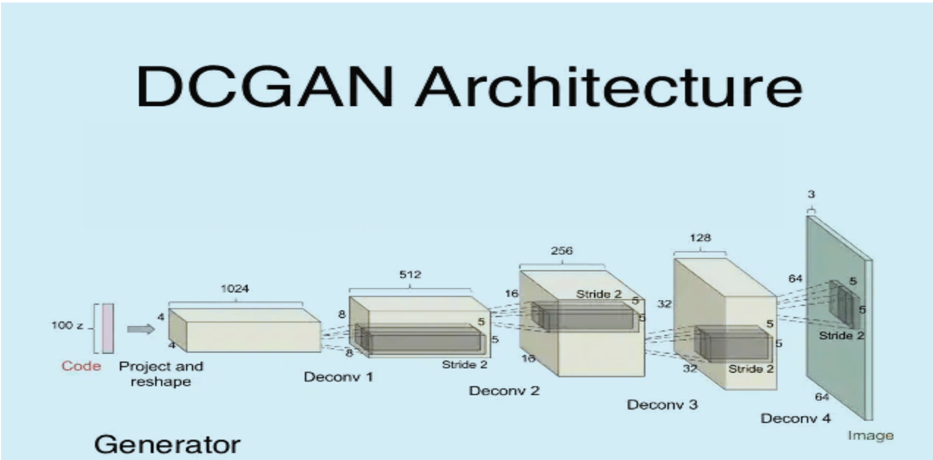

Figure 1.14: DCGAN (Deep Convolutional Generative Adversarial Network)

Autoencoder

An AE is a method for unsupervised feature learning using DNN that efficiently encodes and decodes input. For numerous purposes, including data dimensionality reduction, compression, fusion, and more, autoencoders learn to represent data by encoding it. An auto-encoder system's encoder and decoder are its two primary components. During the encoding process, it is common practice to employ a constructive feature representation to map the input samples to a lower dimensional features space. This process may be carried out again and again until the feature dimensional space that is required is attained. In contrast, feature regeneration from lower-dimensional characteristics via reverse processing occurs during decoding. Figure 1.15 shows the schematic of the auto-encoder, which includes the encoding and decoding stages.

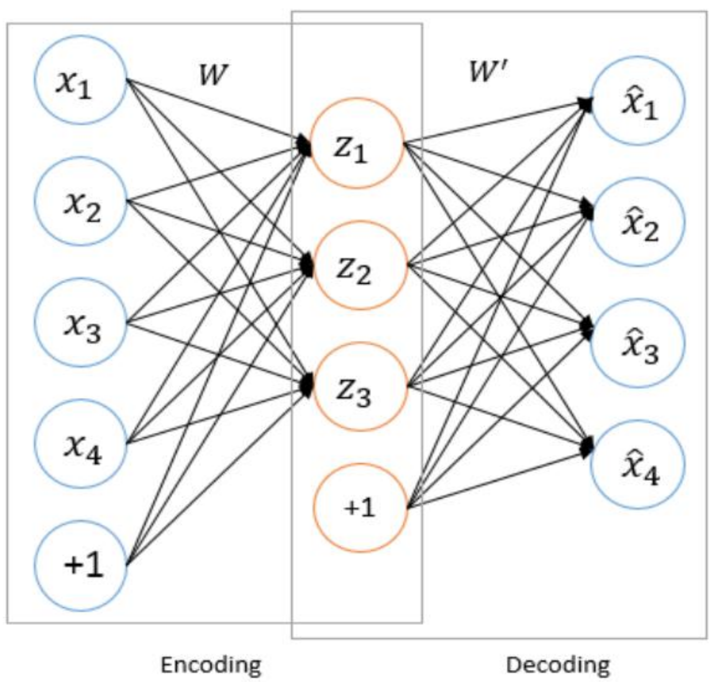

Figure 1.15: Diagram for Auto Encoder.

A transition between the encoder and decoder may be symbolized by the symbols $\emptyset$ and φ.

$$\emptyset : \mathcal{X} \rightarrow \mathcal{F}$$
$$\varphi : \mathcal{F} \rightarrow \mathcal{X}$$

$$\emptyset, \varphi = argmin_{\emptyset,\varphi} \|X - (\emptyset, \varphi)X\|^2$$

.... (1.3)

In the following form, it may be described as an auto encoder with one hidden layer that takes an input of $x \in \mathbb{R}\ d = \mathcal{X}$, and maps it onto $\in \mathbb{R}\ p = \mathscr{F}$:

$$z = \sigma_1(Wx + b)$$

… (1.4)

here W is a weight matrix and b is bias. Element-wise activation functions like sigmoid or ReLU are represented by $\sigma 1$. Let us examine a new mapping or reconstruction of z onto $x\,'$, which has the same dimension as x. One way to represent the reconstruction is as

$$x' = \sigma_2(W'z + b')$$

……. (1.5)

The following is a definition of the loss function that this model is trained with about minimizing reconstruction errors:

$$\mathcal{L}(x, x') = \|x - x'\|^2 = \left\|x - \sigma_2\left(W'\left(\sigma_1(Wx + b)\right) + \ b'\right)\right\|^2 \quad \text{……. (1.6)}$$

In most cases, an input feature space $\mathcal{X}$, which is like a compressed version of the input sample, has more dimensions than the feature space of $\mathscr{F}$. An encoding and decoding stages of a multilayer auto-encoder include the repetition of the same process. The autoencoder's encoder and decoder are extended with numerous hidden layers to form a deep autoencoder. The deeper AE model will continue to have significant issues if the gradient vanishes, or gets too tiny as it continuously traverses several model levels. The parts that follow will go over many sophisticated AE models (Alzubaidi *et al.*, 2021).

Restricted Boltzmann Machine (RBM)

An ANN known as a restricted boltzmann machine allows us to construct non-linear generative models from unlabelled data using unsupervised learning techniques. To teach the network to predict the input, we must increase a function (such the logarithm or product of the vector's probability) in the units that can be seen. The system learns the distribution of probabilities across its inputs. A visible and hidden layer of a network make up RBM, as shown in Figure 1.16. There is a complete absence of connectivity between any two units in either the visible or buried layers (Upadhya and Sastry, 2019).

The following is an expression for an energy(E) function of a visible and hidden unit configuration (v, h):

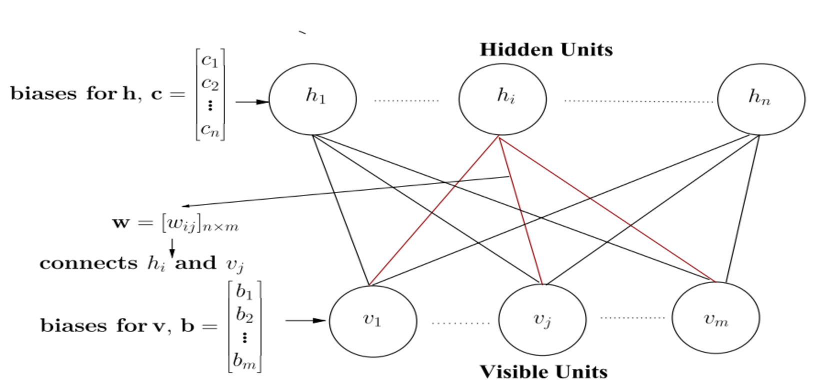

Figure 1.16: RBM with m visible units and n hidden units. wij is the weight between hi and vj and the terms b and c denote the bias for visible and hidden unit, respectively.

Long Short-Term Memory (LSTM)

LSTM models are effective recurrent neural systems that can learn long-term dependencies despite huge minimum time delays, allowing them to circumvent the exploding/vanishing gradient problems. The Constant Error Carousel (CEC) is a general solution to this issue, maintaining an error signal inside every unit's cell. (Hochreiter and Schmidhuber, 1997) An intriguing design that is used to construct memory cells is to enhance the CEC with other features, such as input and output gates. This results in what are basically recurrent networks. There is a one-time step lag in feedback shown by a self-recurrent connection.

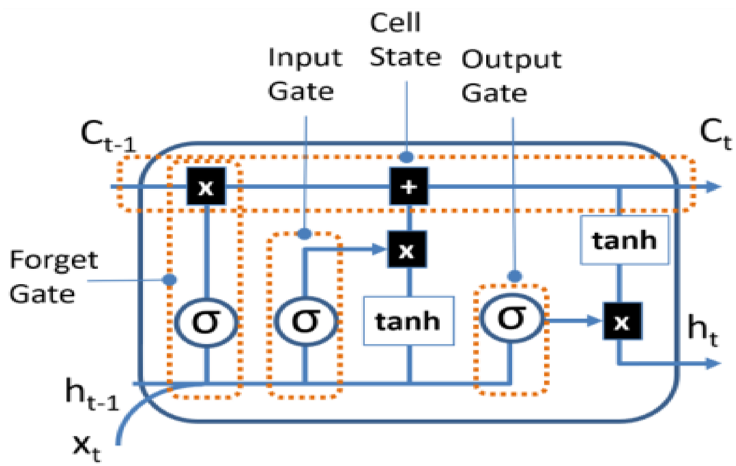

Figure 1.17. LSTM block with memory cell and gates.

Figure 1.17, Cell, input, and output values are represented by C, x, and h, respectively. Subscript t represents the time step value, i.e., $t - 1$ originates by the LSTM block that came before it (or from time $t - 1$), and t indicates the values of the current block. The hyperbolic tangent function is represented as $tanh$, while the sigmoid function is symbolized by σ. Elemental multiplication is represented by the operator x, while element-wise summation is represented by the operator +. These equations explain the calculations of the gates (Mir and sandhu, 2019).

1.19 Problem statement

The Current fake news detection models struggle to accurately differentiate between contextually nuanced content such as satire, opinion pieces, and deliberately misleading information on social media. These models often lack the contextual depth needed to capture subtle semantic cues, tone, and intent embedded within user-generated content. This leads to a high rate of FP and FN, undermining trust in automated content moderation systems.

To address this, **Roberta** is proposed as the core NLP model due to its superior contextual understanding, robust pretraining on larger datasets, and enhanced fine-tuning capabilities compared to BERT, GPT, or XLNet. Unlike BERT, Roberta removes the Next Sentence Prediction (NSP) objective, allowing it to focus better on individual sentence representations. Compared to GPT's unidirectional architecture and XLNet's permutation-based approach, Roberta offers more stable and accurate performance in downstream classification tasks, making it especially suitable for nuanced fake news detection.

1.20 Research question

- How does Roberta enhance the pre-cision and contextual understanding of fake news detection compared to models like BERT, GPT, and XLNet, particularly in distinguishing between misinformation, satire, and opinion-based content?
- What are the limitations of widely used fake news datasets (e.g., PolitiFact, Gossip Cop) in training robust detection models, and how can data augmentation or transfer learning help address issues of bias, language diversity, and label imbalance?
- To what extent does integrating multimodal inputs (e.g., text and associated images or videos) improve the acc-uracy and reliability of fake news classification on social media platforms?

1.21 Project aim and objective

Aim

This project aims to construct a strong system that can identify fake news and social media misinformation using state-of-the-art DL and ML algorithms. The system will also incorporate content enrichment techniques to improve an accuracy and reliability of the detection models.

Objectives

- To enhance data representation by utilizing the PolitiFact and Gossip Cop datasets, which tag real and fake news respectively, to enhance an accuracy and effectiveness of FND methods.
- To improve detection accuracy by leveraging ML and DL algorithms to develop highly reliable models capable of accurately identifying fake news.
- To integrate advanced algorithms by implementing Large Language Models (LLMs) that use distillation techniques, optimizing performance while reducing the number of parameters for more resource-efficient deployments.
- To address misinformation challenges by developing scalable and adaptable detection systems that can effectively tackle the complexity, diversity, and rapid spread of misinformation on social media.
- To ensure that model choices can be understood and to provide clear insights into the elements impacting adaptation, it is important to assess a model's performance employing measures like recall, accuracy, precision, and F1-score.

1.22 Motivation

This research is driven by the pressing issue of false news, which propagates quickly on social media, deceiving people and seriously harming the political, social, and economic spheres. Fake news erodes trust in institutions, influences elections, damages reputations, and poses serious health risks through false information. Its emotional appeal and ability to exploit cognitive biases, such as confirmation bias, make it particularly dangerous, as it spreads faster than factual corrections.

The shortcomings of conventional detection technologies, which are unable to keep up with the volume and speed of false material online, make it more difficult to combat fake news. This work seeks to create efficient methods for identifying and reducing the impact of fake news by using cutting-edge ML and NLP approaches. By protecting the integrity of information and reducing its harmful impact, this research seeks to foster a more informed and trustworthy digital environment.

1.23 Scope of the Research

The pervasive problem of fake news is the primary emphasis of this study because of the far-reaching effects it has in the modern day on society, politics, and the economy. The article covers several types of FN, including satire, parody, fabrication, manipulation, and native advertising. It also examines the impact of cognitive biases such confirmation bias and naïve realism on how the public perceives and allows disinformation to spread. The study further examines the societal consequences of fake news, including its impact on democratic processes, financial markets, public health, and personal reputations, highlighting its ability to erode trust in legitimate media and institutions. Social media's role as a primary medium for the dissemination of FN is critically analyzed, with a focus on the contributions of bots, cyborgs, and human behavior to its propagation.

The study also explores different approaches to detecting false news, including content-based, context-based, and hybrid methods. To improve the accuracy of detection, it uses advanced NLP techniques like pre-processing and word embedding (e.g., Word2Vec, Glove), as well as machine learning models. The study identifies critical challenges in early detection on dynamic platforms like social networks, emphasizing limitations in existing models, such as semantic conflicts and feature correlations. Furthermore, it investigates innovative technologies, including blockchain for verification and crowdsourcing for fact-checking, to develop robust fake news detection systems. The project intends to provide governments, media organizations, and tech developers practical insights to lessen the adverse consequences of false news on society by tackling these difficulties and offering AI-driven solutions.

1.24 Ethical concerns in Automated Fake News Detection

The ethical concerns surrounding automated fake news detection are multifaceted and require careful consideration. One of the primary risks is the possibility of **false positives**, where legitimate content—such as satire, opinion pieces, or dissenting political views—is incorrectly flagged as misinformation. This raises serious concerns about **censorship** and the suppression of free speech, especially in politically sensitive contexts. Additionally, the **lack of transparency** in how AI models make decisions can erode public trust, particularly if users are not informed about why their content was flagged or removed. There is also the issue of **bias in training data**, which can result in unfair targeting of specific groups or perspectives, exacerbating existing inequalities. Finally, there are concerns about the **concentration of power** in the hands of tech platforms that deploy these systems, potentially allowing them to influence public discourse without adequate oversight or accountability. Balancing the **need for effective misinformation control** with the **preservation of democratic values** like freedom of expression requires the development of transparent, explainable, and accountable AI systems, as well as clear regulatory frameworks.

1.25 Organization of the Thesis

This thesis's primary goal is to offer a thorough examination of how to spot and stop misinformation campaigns on social media. The organization is as follows:

Chapter 2: Literature Review

This literature study takes a look at every single strategy for identifying fake news that is currently available, including hybrid, content-based, and context-based approaches. It examines the role of advanced NLP models, ML and DL techniques, such as BERT and Roberta, in tackling the challenges of misinformation. This chapter also identifies gaps in current research and explores the typology of fake news and its implications across various domains.

Chapter 3: Methodology

The methodology that underpinned the study is detailed in this chapter. Word embeddings and multimodal analysis are described as approaches that are used extensively throughout the data collecting, pre-processing, and feature extraction procedures. The chapter further outlines the development of detection models leveraging ML and DL algorithms and discusses the evaluation metrics used to measure model performance.

Chapter 4: Implementation and Analysis

The application of the suggested false news detection methods is the main topic of this chapter. It gives a thorough evaluation of the datasets, including PolitiFact and Gossip Cop, and talks about the technologies and tools that were employed, such Python and Jupyter Notebook. Important results from the experiments are also summarized and the models' performance is assessed in this chapter.

Chapter 5: Results and Discussion

The outcomes of the detection models' validation and testing are detailed in this chapter. It compares the proposed models' performance to that of current approaches and provides a full commentary on their efficacy. Additionally, the chapter delves into the impacts of the results on enhancing systems for detecting fake news and investigates their ramifications.

Chapter 6: Conclusion and Future Work

The last chapter provides a summary of the important findings, focusing on how this study improved the identification of false news. It provides recommendations for future research, including addressing challenges such as scalability, multilingual support, and real-time detection. The chapter concludes by discussing potential directions for further development in this field.

CHAPTER II:

LITERATURE REVIEW

2.1 Background

Information accessibility and distribution rates have rose dramatically since the start of digital technologies. The spread of false information alongside propaganda has grown due to knowledge democratization. News-like false information has become a global issue which influences public emotion and decides election outcomes as well as potentially triggering social instability. The rapid spread of vast amounts of information across various online channels has led to a breakdown in the capacity to distinguish between real and FN (Olan *et al.*, 2024). False news influences true events and creates doubts about traditional media sources in both digital and physical realms. The number of FN reports during the 2016 U.S. presidential election potentially altered voter choices and possibly influenced the election results(Tokpa *et al.*, 2023).

A great deal of misinformation about the COVID-19 virus, its treatments, and vaccinations was also widely circulated throughout the pandemic, which created major hazards to public health(Meesad, 2021). Fast and effective disinformation identification systems prove necessary during these events.

An important tool in the battle against disinformation is the use of word embedding techniques inside NLP. A continuous vector space shows words as vectors in order to capture semantic linkages between concepts through compatible vector representations. Dense vector representations of algorithms have made text input become much easier for algorithms to comprehend and analyze with techniques such as TF-IDF, Word2Vec and Fast Text that has revolutionized everything about text processing. Embeddings function as crucial components for model detection of fine-scale data connections because they improve feature extraction capabilities.

When compared to older, more traditional ML methods, the more recent technique of DL has shown to be more effective in spotting fake news. The following advantages exist specifically within DL, which ML fails to deliver: a) automated feature extraction paired with b) minimal pre-processing requirements alongside c) high-dimensional characteristic extraction and d) superior accuracy levels. The widespread availability of data and programming tools has greatly increased the potency and utility of DL-based approaches. So, many publications on DL-based fake news detection have been published in the last five years. In an eager attempt to compare the vast quantity of research efforts focused on DL-based false news detection, the existing literature has been reviewed.

2.2 Related Work

This section provides a summary of previous research on detecting false news on social media, focusing on several ML and DL methods.

Fake News in Social Media Content Analysis Using Deep Learning

In this study (Akram, Zhou and Dong, 2024) , proposes a new FN classification model based on an improved DL model known as BERT on datasets namely, Politi Tweet and Buzzfeed. The Politi Tweet dataset receives additional data to overcome class imbalance issues by integrating regional linguistic features along with cultural references from different areas. An improved model process detects complex patterns with more accuracy which leads to enhanced classification accuracy. Performance evaluation was conducted against BERT base for assessing both performance and generalizing capabilities of the model. The improved model reached a 98% success rate on Politi Tweet data while delivering a 85% success rate on Buzzfeed content. It displayed better outcomes than baseline prototypes.

In this research (Hashmi *et al.*, 2024), demonstrated a powerful technique for identifying bogus news utilizing 3 publicly available datasets: WELFake, Fake News Net, and Fake News Prediction. By using regularization and hyperparameter optimization they enhanced these techniques that were combined with Fast Text word embeddings so the model could prevent overfitting and demonstrate better generalizability. The combination of CNNs and LSTm with Fast Text embeddings yielded superior results across all datasets achieving F1-scores of 0.99, accuracy of 0.97 and LTM of 0.99. They also made use of cutting-edge transformer-based models that have been fine-tuned by hyperparameter modifications, including BERT, XLNet, and Roberta.

In this study (Al-tarawneh *et al.*, 2024) evaluated fast Text, word2vec and TF-IDF word embedding methods in different ML and DL models for detecting FN effectiveness. The researcher utilized Truth-seeker to test various classifiers while working with news stories and labeled social media postings from more than ten years. The dataset includes SVMs, MLPs, and CNNs. Their research shows that CNNs with TF-IDF embeddings and SVMs with TF-IDF embeddings perform best in every respect, including acc-uracy, pre-cision, re-call, and F1 score. The performance of SVM models that operate efficiently with sparse data representations improves through the use of TF-IDF because it highlights important text discrimination features.

In this research (Baashirah, 2024), proposed a new model called Zero-Shot Learning (ZSL) Fake News Detection (ZS-FND) to solve these problems. The suitability of ZSL models lies in their ability to work without domain-specific data while predicting with minimal training datasets due to their nature of dealing with dynamic and varied false news patterns. Word vectors are produced using the suggested ZS-FND model, which makes use of the pre-trained BERT to represent text effectively. For FND, ZSL uses these vectors as input. Compared to traditional ML and DL models, ZS-FND achieves better results according to accuracy (98.39%), precision (97.33%), recall (95.67%), F1-score (96.49%), and MAE (0.0160). By using ZS-FND, you may increase your accuracy by 10.76%, your precision by 4.05%, your recall by 5.96%, and your F1-score by 5.01%.

In this research (E Almandouh *et al.*, 2024), investigated multiple techniques for Arabic false news detection through ML, DL and combination learning methodologies. Using a variety of ML and DL techniques, they combined Fast Text word embeddings. Then, using sophisticated transformer-based models like BERT, XLNet, and RoBERTa, they carefully adjusted their hyperparameters to maximize their performance. The study's approach consists of applying thorough preprocessing methods to the text data and classifying 2 Arabic News article datasets—the AFND and ARABICFAKETWEETS datasets—into fake and true subsets. CNN-LSTM, RNN-CNN, RNN-LSTM, and Bi-GRU-Bi-LSTM are the four hybrid DL models that are introduced. The F1 score, accuracy, and loss measures all showed better performance with the Bi-GRU-Bi-LSTM model. On the AFND dataset, the hybrid Bi-GRU-Bi-LSTM model achieves precision of 0.97, a recall of 0.97, F1 score of 0.98, and accuracy of 0.98. On the ARABICFAKETWEETS dataset, these same metrics are 0.98, 0.98, 0.99, and 0.99, respectively.

In this research (Tabassum *et al.*, 2024), filled this need by presenting Punny Punctuators' involvement in the Dravidian Language Tech@EACL 2024 joint assignment on Fake News Detection. There are 2 main parts to the collaborative project: 1. identifying authentic or fake social media posts, and 2. 5 categories for classifying fake news. Different processing approaches, such as transliteration, and models based on ML, DL, and transformers were used in their experiments. Malayalam BERT achieved the highest score on both subtasks, placing us in second position with a macro f1-score of 0.87 for subtask-1 and eleventh place with a macro f1-score of 0.17 for subtask-2.

In the study (Balshetwar, Abilash and Dani Jermisha, 2023), advise managing datasets containing multivariate missing variables, such as those found in news or social media data, by using a multiple imputation method that utilizes the Multiple Imputation Chain Equation (MICE). Consequently, TF-IDF is utilized to discover the weighted matrix's long-term characteristics, enabling effective feature extraction from text. To determine the relationship between missing data variables and valuable data properties, classifiers such as DNN, passive-aggressive, and NB are employed. The findings of this research demonstrated that a proposed method's overall computation for identifying false news by analyzing the dataset's evaluation of various claims (e.g., somewhat true, true, somewhat true, and false) attained a precision of 99.8 percent.

In this study, (Rahman, Ashraf and Kabir, 2023) found similar patterns in both real and fake Bangla news items, which may be utilized to detect articles that convey misleading information. They trained and verified their DL model using their selected dataset. 1,299 fake news stories and 48,678 real news pieces make up the learning dataset. They used random under sampling and an ensemble to handle the imbalanced data and get the combined outcome. They achieved a recall of 99% and an acc-uracy of 98.29% using the model they suggested for processing Bangla text.

In this study, (Rasel *et al.*, 2022) gather fresh data and combine it with current datasets in an attempt to create a Bangla false news dataset. They conducted an experiment that included removing duplicate data from existing databases. At last, they build a database of 4,678 unique news records for the purpose of fighting fake news. They achieved advanced results by applying a number of DNN, ML, and Transformer models to their data. Our selection included LSTM, LSTM-CNN, BiLSTM-CNN, BiLSTM, CNN, Adaboost, SVM, DT, and KNN. At 95.5%, 95.9%, and 95.3% accuracy, respectively, CNN-LSTM, CNN, and BiLSTM are the top models. In comparison to previous results when testing with pre-existing datasets, the model accuracy increased by 1.4% to 3.4%. They not only improved accuracy, but their models also significantly outperformed previous research when it came to remembering false news stories.

In this research (Hamed, Ab Aziz and Yaakub, 2023), characteristics that were identified from news articles using sentiment analysis and user comments using emotion analysis. The content feature of the news and these characteristics were inputted into the suggested Bi-LSTM model for the purpose of detecting fake news. The suggested model was trained and tested using the popular Faked it dataset, which includes news headlines and comments made about them. The retrieved attributes allowed the proposed model to outperform prior state-of-the-art research, with an Area Under the ROC curve measure of 96.77 percent.

In this study (GÜLER and GÜNDÜZ, 2023), offered the challenge of detecting social media fake news in two languages, English and Turkish, which had different linguistic characteristics. So far as they are aware, they have created the first publicly available dataset of Turkish fake and true news tweets, SOSYalan. They conduct studies using BuzzFeed and ISOT, two benchmark datasets, to determine the English language. Their DL-based algorithms for detecting false news in English and Turkish are built on CNNs and RNN-LSTMs coupled using the Word2vec embedding model. Metrics like as recall, accuracy, precision, F1-score, and rates of FN, FP, and TR are used to assess the produced systems. The findings demonstrate that the developed methods for the English language attain accuracy rates varying from 85.16 to 99.9 percent, which is in line with the majority of the most recent advances in academic research. Their systems' accuracy ratings for the Turkish language vary from 87.14% to 92.48%, which further proves their superiority compared to the limited research done in this field.

In this research by (Saeed and Al Solami, 2023), the problem is in classifying erroneous information and fake news using data from social media. Two fake news datasets are collected and preprocessed using a number of data enhancement and preprocessing techniques. TF-IDF and BoWs are two n-gram techniques that are used to extract textual characteristics from each dataset separately, whereas Word2vec and Global Vectors for Word representation are used to extract deep text features. Additionally, embedded representations may be derived from the input data by using BERT. Lastly, for the purpose of classifying fake news, three ML algorithms and two DL algorithms are engaged. In the same way as ML and DL models classify embedded outcomes, BERT does the same.

In this study (Tokpa *et al.*, 2023), created two hybrid DL models and tested them on ISOT and FAKES, two real-world datasets. Based on their experience, the presented models were able to achieve a 99% accuracy rate on ISOT and a 68% and 63% accuracy rate on FA-KES, respectively, when the goal was to identify fake news.

In this research (Siram Divya, 2022) put out a model called SPOT, which stands for " detection of false news on social media sites by analyzing user and event trustworthiness and opinion data." This model uses Twitter metadata for credibility analysis and opinion mining on user comments. The suggested approach improves performance by learning about a specific event using the knowledge source that is available during feature extraction. The proposed method considers the text's cognitive cues to allow opinion mining using Senti Word Net. It takes it a notch further by easily detecting false news using a bidirectional Gated Recurrent Neural Network (GRNN) that integrates objective aspects such as emotion and a believability score. The testing findings show that the suggested SPOT method achieves a much greater accuracy rate (14.15% higher) than the current DUAL method.

In this study (Sarnovský, Maslej-kre and Ivancová, 2022), centres on an application of DL models to a task of automatically identifying FN stories written in Slovak. Data utilized to train and evaluate various DL models was sourced from many local online news sources; all of these models were pertinent to the COVID-19 pandemic. By combining a biLSTM network with one-dimensional convolutional layers, scientists were able to attain an average macroF1 score94% on a different training set.

In this research (Kausar, Alikhan and Sattar, 2022), created a hybrid model to identify fake news by enhancing representation learning for better detection of such content. The recommended hybrid model employs a DL model (either LSTM or BERT) to extract sequential features after content-based features have been extracted employing N-gram with TF-IDF. They utilised two publicly available datasets to test how well the suggested technique worked. Based on the results, the proposed strategy is superior to previous methods described in the literature for identifying false news. Through the recommended method, the WELFAKE dataset reached a 96.8% accuracy level while the Kaggle Fake News dataset succeeded at 94%.

In this study (Kanagavalli and Priya, 2022) suggested the RDL-FAFND model is able to identify phoney social media accounts by using a krill herd optimization approach with a parameter adjusted deep stacked auto encoder (DSAE). Additionally, in order to classify the text as real or false, the RDL-FAFND model uses an EML-LF ensemble of ML models with various linguistic properties. Extensive testing has been carried out to verify that the RDL-FAFND model is more effective. The provided RDL-FAFND model outperforms the current approaches, according to a thorough study of the findings.

In this research (Amer, Kwak and El-Sappagh, 2022), tested three different models, classifiers, and transformers using ML and deep learning. They used word embedding to glean article context in every trial. The experimental results demonstrated that DL models achieved higher accuracy than ML classifiers and the BERT transformer. In addition, the LSTM andGRU models were nearly identical in terms of accuracy. The researchers demonstrated success by merging a DL model with better linguistic characteristics which led to accurate fake news detection.

According to (Matheven and Kumar, 2022) this work showcased the creation of a system that can identify false news by utilizing DL and NLP. Researchers incorporated the LSTM model together with Word2vec model when developing the system to prove the two components' compatibility. The currently-in-use system was trained and evaluated using two distinct dataset sets, one of which included actual news stories and the other of which contained fake news. Furthermore, three separate factors were selected to examine their correlation with the system's accuracy: data variety, training cycles, and vector size. Thorough research demonstrates that these three conditions strongly affect a system's accuracy rate. After that, it was trained using the best possible variables and tested to ensure it met the minimal accuracy requirement of 90%. This degree of Accuracy proves that the LSTM and Word2vec models work well together and can be integrated into a single system capable of reliably detecting fake news.

(Jayakody, Mohammad and Halgamuge, 2022) A distributed DL model based on federated learning (FL) is constructed for the purpose of FND. A DL model is trained using the ISOT dataset (N = 44,898) from Reuters.com, which contains fake news stories. Then, recall, precision, F1-score, and accuracy are utilized to compare the two types of centralized and decentralized models. Moreover, by varying the number of FL customers, the efficacy was evaluated. Without the usage of pre-trained word embeddings, they are able to identify a remarkable accuracy of their recommended decentralized FL technique (99.6% accuracy) with less communication cycles than earlier research. The most accurate comparison is to three previous research articles that have used the same technique. Instead of using a centralized system to detect fake news, the FL technique may be employed more effectively. Improving the trustworthiness and veracity of news sources is possible with Blockchain and comparable technologies.

(R. H. Khan *et al.*, 2022) aims to optimize a model that can identify FN using a flexible and resilient feature extraction approach. Using a well-known dataset of fake news that can be found on Kaggle, this research is completed. The method that has been proposed uses what is formally known as "Stemming" to break down each expression into its component words. Both TF-IDF and BERT transform every word of text into a feature vector before sending it on to ML and Deep Learning (BERT). The performance analysis found that out of all the strategies that were investigated, BERT paired with stemming NLP methodology produced the greatest accuracy (99.74%). Just 98.90% accuracy was achieved by the prior gold standard approach. This speed gain is mostly attributable to stemming, which simplifies a phrase by reducing it to its root word. This process produces a more general vector, which, in turn, enhances the model's accuracy. While the SVM (linear kernel) corrected 98.99% of the errors, the passive-aggressive classifier method using a stemmed TF-IDF vectorizer had a higher accuracy level of 99.11%.

(Guo, Lamaazi and Mizouni, 2022) suggested that this study offers a framework for FND in many regions. In a mobile crowdsensing (MCS) context, the framework is utilised to choose reporters based on their accessibility in a certain region. The selected individuals disseminate the news to the nearest edge node, where it is processed locally and false news is detected. With a 91% success rate, the pre-trained BERT model is a powerful tool for detection.

(Abbas, Zeshan and Asif, 2022) suggested method assigns words certain vectors based on the outcomes of the preprocessing stage. A word's inherent characteristics are represented by each vector that is allocated to it. The LSTM model is employed after the resultant word vectors have been inputted into RNN models. The news article's credibility is assessed using the LSTM's output.

(Wang, 2022) claimed that after comparing the GCN method to other neural network models, it was shown to have an appropriate accuracy of around 85% in recognising fake news. Due to the absence of a standard training dataset, it also suggests that future models for detecting fake news should emphasize unsupervised and semi-supervised learning.

(Mahara and Gangele, 2022) Presented deep learning models for Bi-LSTM and LSTM to detect fake news. The initial step in text cleaning with the NLTK toolbox is to remove stop words, special characters, and punctuation. The text undergoes tokenization and preprocessing using a comparable suite of tools. In subsequent iterations, GLOVE word embeddings have included preprocessed text with word sequences learnt by the Bi-LSTM and RNN-LSTM models. The suggested approach improves efficiency by using thick layers and extra dropout techniques. The suggested RNN Bi-LSTM-based method uses the Adam optimizer to a 94% success rate, whilst the Binary cross-entropy loss function with a Dropout loss yields a 93% success rate. As the dropout range grows, the model's accuracy diminishes; using a dropout of 0.3 results in an accuracy of 92%.

(Kumar, Singh and Singh, 2022) this study presents a DL ensemble model that can detect COVID-19-related Twitter misinformation. Different algorithms have been trained to differentiate among real and fake data using text data connected to COVID-19:

CT-BERT, Roberta, and BER Tweet. They compared many classic DL and ML models to your proposed ensemble-based method. The proposed ensemble-based DL model detected COVID-19 FN with a weighted F1 score of 0.99 when applied to Twitter data.

(Garg and Sharma, 2022) this study examined CNN's usage of DL classifiers to detect deceptive news. Using the Reuter dataset and the proposed approach, experiments are carried out. An improvement over the current methodologies, the proposed approach attained an accuracy of 93.64 percent.

(Collen, Nyandoro and Zvarevashe, 2022) said that creating a model for FND by analyzing headlines and supplementary material is the goal of this study. Prior scholarly work suggests that the models' underwhelming performance was brought about by faulty text categorization fine-tuning and the lack of feature extraction. This study will find the difference by using a 5L-CNN DL model that has word embedding tokenizers built in and is more accurate than traditional ML models. This research compares ML algorithms for news article credibility classification with DL algorithms (RNN, LSTM, and 5L-CNN). With a success rate of 99.99%, the model presented in this research achieved the maximum accuracy possible with 5L-CNN.

(Hangloo and Arora, 2022) focused on methods for Content-Based Fake News Detection (FND) that make use of both textual and visual data. The Multimodal technique has been used more recently; it combines visual and textual data. This article delves into many strategies for content-based FND using a DL methodology. Using multimodal approaches considerably improves overall performance, according to the outcomes of a thorough evaluation of various deep learning frameworks and models. In addition, there are a number of potential future paths that might help improve the efficiency of FND frameworks. In addition to outlining possible answers to the problems that may arise when building a FND framework, this study lists a variety of problems that researchers may encounter.

In this study (Ahmad *et al.*, 2022), concentrated on the difficulties caused by multiple rumours that spread quickly through social media networks rather than persistent rumours. They proposed new features that take social media and content into account in order to detect rumours on social media. In terms of rumour categorization, their proposed features outperform the state-of-the-art baseline characteristics. In addition, they use text-based bidirectional LSTM-RNN for rumour prediction. The rumour detection methodology is straightforward, yet it works. Most of the first work on rumour detection assumed that rumours were always unfounded and concentrated on persistent rumours. On the other hand, they use a real-world scenario data set for their rumour detection tests.

In this research (News *et al.*, 2022), developed Mc-DNN, a multi-channel DL network that uses various channels to evaluate news headlines and articles in order to distinguish among authentic and fake news. With ISOT Fake News Dataset, they reach a maximum accuracy of 99.23%, while with Mc-DNN, they reach an accuracy of 94.68%. As a result, they strongly advise using Mc-DNN for detecting fake news.

In this research (Fouad, Sabbeh and Medhat, 2022), offered a model architecture that relies solely on textual elements to identify Arabic fake news. Data mining and other forms of deep learning were employed. A number of DL models are used, such as CNNs, LSTMs, and BiLSTMs. The findings show that when it comes to accuracy rate, the BiLSTM model is the best model while training employing recursive and basic data split modes.

In this study (Aslam *et al.*, 2021) Employing the LIAR dataset, a DL model was developed that used an ensemble technique to assess the veracity of news articles. The dataset was optimized using two DL models. The "statement" attribute was trained using a Bi-LSTM-GRU-dense model, whereas all other attributes were trained using a dense DL model. Results from experiments using just the statement attribute showed that the proposed technique obtained 0.898 accuracy, 0.916 recall, 0.913 precision, and 0.914 F-score.

In this study (Kaliyar, Goswami and Narang, 2021), suggested a DL method called Fake BERT that is based on BERT. This method integrates BERT with many DCNN blocks running in parallel, each with a unique filter and kernel size. The biggest obstacle to natural language processing is ambiguity. However, this combination can help with that. Classification results show that their suggested Fake BERT model achieves a 98.90% accuracy rate, which is better than the advanced models.

In this study (Wani *et al.*, 2021), examined data mining-based automated methods for detecting FN. Applying the Contraint@AAAI 2021Covid-19 FND dataset, they assess several supervised text classification methods. BERT, LSTM, and CNN constitute the basis of the classification techniques. They also assess the significance of dispersed word representations and language model pre-training using the unlabelled COVID-19 tweets corpus as examples of unsupervised learning. Results on the COVID-19 FND Dataset show an impressive 98.41% accuracy.

In this research (Martínez-Gallego, Álvarez-Ortiz and Arias-Londoño, 2021), investigated various training approaches and designs to provide the groundwork for future studies in this field. The Deep Learning models were constructed using several pre-trained word embedding representations, like GloVe, ELMo, BERT, and BETO (a Spanish variant of BERT). The findings showed that a RNN with LSTM layers and a pre-trained BETO model worked well, with an accuracy of up to 80%; however, a model utilizing a RFestimator as a baseline model also achieved comparable results.

In this research (Saleh, Alharbi and Alsamhi, 2021) Discovering the best model that achieves excellent accuracy performance is the objective. So, they came up with the idea of an OPCNN-FAKE model. Utilising four benchmark datasets for fake news, they assess OPCNN-FAKE in comparison to RNN, LSTM, and six traditional ML methods: DT, LR, KNN, RF, SVM, and NB. Hyperopt optimization was used for ML parameters and grid search for DL parameters.

In this research (Hossain *et al.*, 2021) developed a corpus to identify fake news articles in Bangla by training it with 57,000 Bangla news items on authenticity and counterfeiting. In this training, the Bi-LSTM with Glove model achieved 95% accuracy and the Fast Text model 94% accuracy by implementing K-fold cross-validation. In this study the Gated Recurrent Unit (GRU) produced 77% accuracy when applied as a modern analysis method. By way of comparison, they were able to track down an accuracy of 96% employing the Bi-LSTM, which is consistent with their suggested model.

In this study (Agarwal *et al.*, 2020), explored an use of GloVe for text pre-processing, with the goal of building a word vector space and establishing a linguistic connection. Benchmark results in FND were accomplished by the suggested model, which is a combination of CNN and RNNs design. The model was further enhanced with the usefulness of word embeddings. Additionally, several model parameters have been fine-tuned and documented to guarantee high-quality predictions. Among other variants, including a dropout layer lowers the model's overfitting and produces noticeably improved accuracy metrics. When compared to other solutions for this problem, such as feed-forward networks, RNN, or GRU, it produces greater precision values (97.21 percent) while taking into account more input data.

In this study (Jiang *et al.*, 2020), made many recommendations for ML and DL models. However, the vast majority of them fail to accurately detect fake news. As a result, they suggested a DL architecture that has a 99.82% accuracy rate in differentiating among fake and authentic news. A dataset often utilized for fact-checking was utilized for training and evaluation of this BiLSTM model. Execution time, precision, recall, and F1-measure were among the model evaluation metrics they used to prove their model's efficacy.

In this study (Pre-proofs, 2020), found that it was essential to a technique for recognizing and categorizing fake news. The substance is what really makes a fake news story convincing to its target audience. As a result, we suggest a linguistic model to identify the traits caused by language in the text. This language model analyzes specific news articles for their syntax, grammar, sentiment, and readability. When dealing with the curse of dimensionality, language-driven models necessitate a method to manage features that are both labor-intensive and individually designed. Because of its superior performance in identifying fake news, they use a sequential learning model based on NN. The combined linguistic feature-driven model achieved an average acc-uracy of 86% in identifying and classifying FN, proving that the linguistic model's characteristics were important.

In this research (Umer *et al.*, 2020), suggested using dimensionality reduction methods to make the feature vectors more manageable before feeding them into the classifier. This study utilized a dataset with four different positions—agree, disagree, discuss, and unrelated—to construct its reasoning from the Fake News Challenges (FNC) website. The contextual characteristics used to detect FN are improved with the addition of nonlinear features to PCA and chi-square. This study was motivated by a desire to understand the sentiment behind news headlines. The accuracy is improved by 4% and the F1 − score is improved by 20% using the suggested model. In the studies, PCA outperformed both Chi-square and advanced methods with an accuracy of 97.8 percent.

In this research (Saikh *et al.*, 2020), put out two deep learning-based models that effectively address the issue of FND in various online news contents. Two newly available datasets, Fake News AMT and Celebrity, are used to test their methodologies for fake news identification. The suggested approaches do better than a present advanced system, which is based on handmade feature engineering, by a considerable margin of 3.08% and 9.3%, respectively. They are encouraging. They do cross-domain analysis to see how well their systems work in different domains, so they may make use of the datasets that are available for similar tasks. For instance, they may test a model on Celebrity after training it on Fake News AMT, and vice versa.

In this study (Kaliyar *et al.*, 2020), presented FNDNet, a DCNN for FND. As an alternative to relying on manually created features, they trained their model (FNDNet), a DNN with several hidden layers, to automatically learn a differentiating quality for fake news classification. To do this, they build a multi-layer deep CNN. They evaluate the suggested method in comparison to many baseline models. A proposed model achieved advanced performance after training and testing on benchmark datasets, achieving a test data accuracy of 98.36%. Precision, recall, F1, accuracy, FP, TN, Wilcoxon, and other performance evaluation metrics were used to verify the findings.

In this study (Kumar *et al.*, 2020), have gathered 1356 news cases from different individuals using Twitter and media sites like PolitiFact, and have created many datasets for both actual and fake news articles. The researchers examine attention mechanisms together with ensemble approaches as well as CNNs and LSTMs in their evaluation process. The researchers found that the network with the greatest accuracy (88.78%) was the CNN + bidirectional LSTM ensembled network with attention mechanism.

In this research (Abdulrahman and Baykara, 2020), concentrated on textual content categorization in order to identify and categorize social media fake news. Using ten distinct ML and DL classifiers, this classification employed 4 conventional methods to extract text features (TF-IDF, Count Vector, Character Level Vector, and N-Gram level vector) in order to classify the fake news dataset. Outcomes shown that textual fake news may be identified, particularly with the use of a CNN. From 81% to 100% accuracy was achieved throughout this training utilizing several classifiers.

In this study (Abedalla, Al-Sadi and Abdullah, 2019), meant to illuminate the problem of FN and the procedures for detecting it using DL methods. Using the Fake News Challenge (FNC-1) dataset, they have developed a number of algorithms to analyse the link among the article's title and content in order to detect fake news. Building blocks for their models mostly consist of CNNs, LSTMs, and Bidirectional LSTMs. Their research obtained 71.2% accuracy for the official testing dataset, which is in contrast to other studies on the same dataset that claimed accuracy for test data drawn from the same training dataset.

In this research (Qawasmeh, Tawalbeh and Abdullah, 2019), studied the automatic FND on internet communication networks. Furthermore, they proposed using modern ML methods for automated FND. An implementation of the proposed bidirectional LSTM concatenated model generated 85.3% accuracy measurement on the FNC-1dataset.

In this research (Monti *et al.*, 2019), presented a novel approach to automatically detecting fake news using geometric DL. Standard CNNs can be fundamental to analyze data collections including content pieces with user behavioral information and social networks and news segments. Professional organizations used fact-checked news stories shared on Twitter to train and test their algorithm. The research demonstrates high accuracy at 92.7% ROC AUC for identifying FN through their evaluation of social network structure and propagation as essential factors.

In this research (Thota, 2018), the issue of FND was resolved by using DL architectures. An exponential development in the creation and spread of misleading information has made the urgent need for automatic tagging and identification of such biassed news items all the more paramount. Nevertheless, automated fake news identification is challenging since it necessitates the model to comprehend subtleties in natural language. In addition, most current models for detecting FN see the issue as a simple binary classification, which restricts the model's capacity to ascertain the degree to which the reported news differs or is identical to actual news. To fill these shortcomings, they provide a neural network design that can correctly anticipate the relationship between a title and the substance of an article. On test data, their model achieves an accuracy of 94.21%, which is a 2.5 percentage point improvement over previous model designs.

In this study (Girgis, Amer and Gadallah, 2018), they want to create a classifier that can tell the difference between legitimate and fake news articles just based on the content. They plan to tackle this problem using RNN method models (vanilla, GRU) and LSTMs, which are deep learning techniques. By applying them to the dataset they used, LAIR, they will demonstrate the difference and analyse the findings. Results are close, however GRU (with a score of 0.217), LSTM (with a score of 0.2166), and vanilla (with a score of 0.215) are the best. Because of this, they think that the accuracy of outcomes will be much improved when GRU and CNN algorithms are used together to the same dataset.

In this research (Singhania, Fernandez and Rao, 2017), used a 3HAN in tandem with an automated detector based on DL to identify FN quickly and accurately. By starting with the most basic building blocks and working its way up to more complex ones, 3HAN is able to efficiently generate news vectors, which are representations of input news stories. Words, phrases, and the headline all have their own level. A hallmark of FN is an exaggerated or misleading title, and just a small number of words or phrases really have a significant impact on the reader. 3HAN's three levels of attention allow it to assign varying degrees of priority to different sections of an item. Their trials on a big real-world dataset show that 3HAN is successful with an accuracy of 96.77%.

Fake News in Social Media Content Analysis Using NLP

In this research (Chang, 2024), contrasted the accuracy with which various systems identified fake news. The training dataset was developed using the ISOTdataset, which includes 21,417 real news items and 23,481 FN stories. Next, they made sure that both the LSTMandBi-LSTM models included the following layers: input, embedding, dropout, LSTM, and output. The train data was used for 10 epochs by both the LSTM and BiLSTM models during training. The test data was also used to assess the trained models. By splitting the input text into two datasets, the BERT deep learning model was finally ready for testing. After restrict the length of the input sequence, the train and test datasets were tokenised and padding was used. Following that, the BERT model underwent 10 epochs of training using the train data. BERT performed evaluations using the test dataset after its model training process was complete. BERT demonstrated the highest accuracy rate of 99.95% while Bi-LSTM achieved 99.00% accuracy when the model results were evaluated in comparative testing.

In this study (Madani, Motameni and Roshani, 2024), proposed using two successive stages of NLP and ML to develop their algorithm. The initial step is to take new samples and extract two structural traits along with other important components. Step two involves enhancing DL model performance using a hybrid approach based on curricular strategies, which makes use of statistical data and a KNN algorithm. The suggested model displayed superior performance to benchmark methods for detecting fake news according to the experimental outcomes.

In this research (Mahmud *et al.*, 2024), analyzed whether fake news detection systems could be automated through the combination of DL and NLP. The researchers studied multiple neural network structures that included CNN, LSTM, Bidirectional LSTM, ANN and the hybrid design of CNN+Bidirectional LSTM. Rigorous testing conducted by their team revealed the CNN and Bidirectional LSTM model as the most efficient solution because it reached 98.13 percent accuracy. The model effectively unites text pattern recognition convolutional layers with bidirectional LSTM layers that track temporal dependences to enhance its abilities for detecting fine text indicators of fake news.

In this study (Prabhakar, 2024), implemented an original ensemble stacking method specifically for FND. For this method, the LIARdataset is utilised. In order to make the dataset more accurate and efficient, it is cleaned and pre-processed. To encode the text into numbers, concepts from NLP like BoW and TF-IDF were adopted. In order to accommodate and represent multivariate data which is complex and often four or more dimensioned, two important machine learning algorithms were used namely the t-SNE and PCA. A total of seven classifiers—two each from ML and DL—and two hybrids of the two types—are given the pre-processed data. The analysis used Adaptive Boosting, Decision Tree, Naive Bayes, Random Forest and Logistic Regression ML methods. A MLP model, which is a kind of DL, was employed. Both MLP and t-SNE were utilised in conjunction with PCA. The implementation used f1-score together with support and recall and precision methods to evaluate and rate each model's performance.

In this study (Veeraiah *et al.*, 2024), uses TensorFlow deep learning framework together with NLP technology to address the FN problem. An use of NLP techniques has made computers quite good at analysing text for biases and linguistic patterns that are characteristic of fake news items. Users can develop complex DL models within TensorFlow for identifying FN in IoT devices precisely. The team aims to establish an IoT system which detects biased expressions along with exaggerated statements and deceptive content in fake news. They will use NLP and TensorFlow for this purpose.

In this study (Farooq *et al.*, 2023), looked into these matters and suggested ways to make Urdu news more classy in terms of fake news. There are 4097 news items in the collection, which covers nine distinct areas. Combining n-grams with a BoW and the TF-IDF allows for experimental results. The key innovation here is feature stacking, which merges verbs taken from preprocessed text with the feature vectors from that same text. Ensemble models like RF and ET, as well as SVMs) and KNN, were used for bagging. Then, in the stacking phase, LR was employed as the meta learner and RF and ET as the base learners. They used five-fold and independent set testing to make sure our models were solid. Stacking obtains scores of 93.39% for accuracy, 88.96% for specificity, 96.33% for sensitivity, 86.2% for MCC, 93.17% for ROC, and 93.17% for F1 score, according to the experimental data.

In this study (Alawadh *et al.*, 2023), drawn from a newly released collection of expert-annotated Arabic fake news. In addition, ML classifiers are fed embeddings based on the Arabic language, while DL classifiers are fed feelings derived from the Arabic language using Arabic-language-based trained miniBERT. Deep neural classifiers based on mini-BERT as well as ML classifiers are provided with the holdout validation approaches. An increase in the amount of training data consistently leads to better performance from mini-BERT-based classifiers, which beat ML classifiers.

In this research (Alarfaj and Khan, 2023), investigated distinct ML and DL methods for classifying FN. They utilized a renowned "Fake News" dataset that had been acquired by Kaggle and included a set of news items that had been annotated. A number of ML models were used, including LR, PAC, MNB, GNB, and BNB. The researchers examined DL models which included CNN, CNN-LSTM and LSTM in their investigations. Important assessment criteria including recall, accuracy, precision, and the F1 score were utilized to compare a models' performance. Along with hyperparameter tweaking and cross-validation they used the techniques to validate the system was functioning correctly. The research outcomes reveal clear advantages and limitations of fake news detection methods with specific examples of their operational weaknesses. The authors demonstrated superior performance of DL models and specifically LSTM and CNN-LSTM models versus traditional ML models. The models demonstrated superior accuracy together with robustness during classification duties. DL models provide effective solutions against fake news dissemination while demonstrating the necessity for complex approaches to solve such complicated problems.

In this study (Altheneyan and Alhadlaq, 2023), derived from the FNC-1 dataset where four categories of misinformation are outlined. Using big data technologies (Spark) and ML, the most advanced techniques for identifying fake news are examined and contrasted. This method constructed a multi-layer ensemble model using a dispersed Spark cluster. Count vectorizer, Hashing TF-IDF, and N-grams were the three feature extraction techniques employed in the proposed stacked ensemble classification model. The suggested model demonstrates superior classification performance than the baseline method through its 92.4% F1 score while the baseline score remains at 83.1%.

In this research (Hosseini *et al.*, 2023), place an emphasis on identifying textual content, such as fake news, by means of interpretable features and procedures. They used a Bayesian admixture model—specifically, a dense representation of textual news—to extract semantic topic-related features, and their deep probabilistic model included a variational autoencoder and bidirectional LSTM networks. Extensive experimental trials employing three real-world datasets demonstrate that their approach aids in model interpretability from the taught themes and performs comparably to advanced competing models.

In this study (Yadav *et al.*, 2023), examined several DL and ML methods for automatically FND in news headlines and descriptions. Word2Vec, GloVe, and FastText were among the word-embedding approaches utilised to provide data representations that were both effective and efficient. They used CNN-LSTM, CNN-Bi-LSTM, LSTM, and BiLSTM, among other deep learning models, for categorisation. By combining two publicly accessible datasets, allData and false and Real News, a huge dataset was created to address the lack of a standardised, vast dataset for detecting false news. The collection contains 64,934 news stories that have been labelled. When applied to this combined dataset, the Word2Vec word embedding method outperformed CNNBiLSTM with regard to acc-uracy (0.975), pre-cision (0.984), re-call (0.970), F1 measure (0.977), and AUC (0.992).

In this research (Nadeem *et al.*, 2023), improved the identification of propaganda and FN by incorporating the idea of symmetry into DL approaches for enhanced NLP. A hybrid HyproBert model is proposed in this work for the automated identification of FN The suggested HyproBert model starts by embedding words and tokenising them using DistilBERT. The spatial characteristics are highlighted and extracted using the embeddings, which are sent into the convolution layer. Afterwards, BiGRU is given the output in order to extract the contextual characteristics. Together with the self-attention layer, CapsNet models the hierarchical link among the spatial features by going to the BiGRU output. The last step in classifying features is to apply a thick layer. The suggested HyproBert model is evaluated using two datasets, ISOT and FA-KES, which include fake news. This allowed HyproBert to outperform both baseline and advanced models.

In this study (Jawad and Obaid, 2023), suggested model was tested using the FNC-1 dataset. A competitive dataset is seen as a global open problem and challenge. The method by which this system operates entails using several NLP algorithms to the content found in the body text and banner columns. Afterwards, the elbow truncated approach is used to minimise the retrieved features, and the soft cosine similarity method is utilized to identify the similarity among each pair. Deep learning methods like CNN and DNN include the new functionality. Aside from the disagree category, the suggested approach accurately identifies all the others. This allows the system to get an accuracy of up to 84.6%, placing it in second place according to previous competing research conducted on this dataset.

In this research (Truică and Apostol, 2023), presented a novel method that reliably classifies news items as either fake or reliable by using document embeddings to construct several models. Additionally, they give a benchmark for several architectures that use binary or multi-labeled classification to identify fake news. They used accuracy, precision, and recall to assess the models on five big news corpora. They outperformed the most advanced Deep Neural Network models, which are notoriously difficult. Rather than the complexity of the classification model, they find that document encoding is the most critical component for achieving high accuracy.

In this study (Mallick, Mishra and Senapati, 2023), developed a method to identify FN stories by employing cooperative DL.The proposed method ranks news stories according on user ratings and comments that measure their trustworthiness. Content with a higher ranking is acknowledged as authentic news, whereas lower-ranked material is retained for the purpose of language processing to guarantee its authenticity.

Converting user input into rankings is done within the DL layer using CNN. Just so the CNN model may learn from bad news, it is fed back into the system. The proposed approach outperforms all other language processing-based algorithms with an acc-uracy rate of 98% in identifying fake news.

In this research (Jaybhaye *et al.*, 2023), provided an extensive analysis of methods based on ML andDL for identifying FN. The results of this study might be helpful for those working on algorithms to detect false news employing ML and DL techniques. It is common practice for news reporters to check the credibility of news reports before airing or publishing them. Using fake news detection techniques to weed out false news allows reporters to maintain concentrate on providing valid and accurate information.

In this study (Shaik *et al.*, 2023), proposed a ML-based method for detecting FN. A classification model that evaluates the visual and linguistic content of news items is generated by using four types of ML methods. This is the accepted technique. The simulation results show that the proposed model outperforms many existing algorithms when applied to a large dataset of actual and fake news articles. Some ML techniques that have been suggested as possible answers to the problems with anti-fake news detection include LR, DT, RF, and passive aggressive algorithms.

In this study (Rasel *et al.*, 2022), attempted to compile a Bangla FakeNews dataset by integrating secondary datasets with freshly acquired fake news data. They eliminated unnecessary data from previously accessible datasets during their trial. Last but not least, they compile 4,678 unique pieces of news data into a Fake news dataset. Applying several Machine Learning (LR, SVM, KNN, MNB, Adaboost, and DT), DeepNeuralNetwork (LSTM, BiLSTM, CNN, LSTM-CNN, BiLSTM-CNN), and Transformer (Bangla BERT, m-BERT) models to their data allowed them to get state-of-the-art outcomes. Three top models—CNN, CNN-LSTM, and BiLSTM—are achieving

95.9%, 95.5%, and 95.3% accuracy, respectively. Using the pre-existing datasets, they also evaluated their models and found an accuracy boost of 1.4% to 3.4% compared to the previous findings.

In this study (Segura-Bedmar and Alonso-Bartolome, 2022), they use both unimodal and multimodal techniques to classify FN in a fine-grained manner on the Fakeddit dataset. According to their tests, the multimodal strategy that combines text and visual data using a CNN architecture yields the best outcomes, with an accuracy87%. The use of graphics greatly enhances some types of fake news, including satire, false connections, and manipulated material. The use of pictures also enhances the outcomes in the other categories, but to a lesser extent. When it comes to text-only unimodal techniques, the most effective model is BERT, which achieves an accuracy of 78%.

In this research (Palani, Elango and Viswanathan K, 2022), created a multimodal feature vector with high information richness by merging textual and visual data; this system can detect fake news automatically. While extracting textual characteristics, the suggested approach makes use of the BERT model, which maintains the semantic connections between words. The suggested Caps Net model separates out the most useful visual characteristics from images, in contrast to the CNN. Combining these factors results in a more complete picture of the facts that may help determine if the news is fake or real. Using two open-source datasets, PolitiFact and Gossip cop, they tested their model's accuracy against several benchmarks. Classification accuracy on the PolitiFact dataset is 93% and on the Gossip cop dataset it is 92%, respectively, which is a considerable improvement above the Spot Fake+ model's 84.6% and 85.6%, respectively.

In this study (Rafique *et al.*, 2022), investigated the development and performance of multiple ML classifiers and a DL model to detect FN in Urdu. To achieve this goal, many methods such as LR, SVM, RF, NB, gradient boosting, and passive aggressiveness have been employed. Terms, inverse document frequencies, and BoW characteristics have all been studied for their potential impacts. A dataset consisting of 900 news items that were gathered by hand was utilised for the tests. According to the results, when it comes to Urdu fake news with BoW characteristics, RF performs the best and reaches the greatest accuracy of 0.92.

In this study (Ouassil *et al.*, 2022), a hybrid CNN and BILSTM model is combined with a number of word embedding methods to create a revolutionary DL strategy for identifying FN. Using the objective dataset WELFake, they trained the model for classification. A CNN on BILSTM layer synergy between two Word2Vec models, one with a pre-trained CBOW model and the other with a Word2VecSkip-Word model, yielded the best results. By working together, these factors were able to attain an accuracy of 97%.

In this study (Rahmawati, Alamsyah and Romadhony, 2022), sought to identify false news stories by utilizing IndoBERT, SVM, and NB to categorize Indonesian news and identify the most effective model. News websites like Detik, Liputan 6, Kompas, Cek Fakta, and Turnbackhoax are included in the gathered data. There were a grand total of 2000 news items, 1000 of which were fake news stories and 1000 of which were real. The IndoBERT algorithm achieves the highest accuracy at 90%.

In this research (Prachi *et al.*, 2022), proposed feature vectors are created using a variety of feature engineering techniques, including regular expressions, tokenization, stop words, lemmatisation, and TF-IDF. Several metrics, such as recall, accuracy, precision, F-1 score, and the ROC curve, were used to assess all of the models in the ML and NNLP disciplines. The classification accuracies for the ML models were as follows: 73.75% for LR, 89.66% for DT, 74.19% for naive bayes, and 76.65% for SVM. In conclusion, the LSTM achieved 95% accuracy, with the NLP-based BERT approach achieving the maximum accuracy of 98%.

In this study (Lahby and Yassine, 2022), given international news source significantly affects their society and culture, for better or bad. The company's reputation runs the danger of being damaged as an outcome of the widespread use of FN and misleading content on social media. False news has spread like wildfire in this type of epidemic, adding to the general public's understandable state of terror. The chapter delves into the development of a system that utilises RNN and its approaches, such as LSTM and BiLSTM, to identify deceptive information. It employs MachineLearning and NLP to address an issue. FN in general and Covid-19 specific fake news have both been implemented.

In this research (Isa, Nico and Permana, 2022), offered IndoBERT, a transformer model based on Indonesian BERT that prioritises the input phrases' context and attention. In preparation for the experiment, they used the data acquired from turnbackhoax.id to fine-tune the suggested model and optimise its hyperparameters. They then ran tests to see how well the model worked, and it got a 94.66% on all three metrics (precision, recall, and F1-score).

In this research (Dhar, 2022), objective of the initiative is to identify authentic news stories that are available online. More and more people are turning to internet media for their news because of the many advantages it offers over traditional print publications, such as being paperless, having the news at the touch of a finger, and having access to attention-grabbing headlines and colourful titles. Recent developments in online social media have an effect on people's day-to-day lives. It is imperative that users of the aforementioned online news platform find a difficult solution that can recognise fake news so that they do not receive it. Using ML and DL, the project's application is to identify fake news. Adaptive Boosting Classifier (ABC) and LSTM are the two main algorithms used in the research. There is also a Passive Aggressive Classifier. They have done excellent work analysing and visualising the provided dataset. The accuracy that their models are producing is 95% or higher. They have completed the development of the web application that can forecast both fake and real news.

In this research (Meesad, 2021), outlined a system for spotting fake news in Thailand. The three primary components of the framework are ML, NLP, and information retrieval. The first part of this study is gathering data, and the second part is creating a ML model. They used NLP techniques to extract valuable qualities from data collected by Thai online news websites utilizing web-crawler information retrieval. Naïve Bayesian, LR, Multilayer Perceptron, SVM, DT, Random Forest, Rule-Based Classifier, and LSTM were among the popular ML models chosen for comparison. After determining that LSTM performed best on the test set, they put their automated web application to work detecting fake news.

In this study (Uma Sharma, Sidarth Saran, 2021), sought to implement a system for binary categorisation of different internet news items using principles from AI, NLP, and ML. Their goal is to let users mark news stories as legitimate or fake and verify the website's legitimacy.

In this research (Hansrajh, Adeliyi and Wing, 2021), recommended using a publically accessible dataset to train a mixed-model machine learning ensemble model that incorporates LR, SVM, LDA, SGD, and ridge regression to determine the veracity of news reports. They will evaluate the proposed model using well-known classical machine learning methods and then quantify its performance using f1-score, AUC, recall, accuracy, precision, and ROC. The suggested model surpassed other well-known classical machine learning models, according on the results given.

In this study (Shahbazi and Byun, 2021), put forward a method that integrates blockchain with NLP in order to employ ML to identify false news and improve the prediction of fake user accounts and postings. To do this, the Reinforcement Learning method is utilised. The decentralization of blockchain infrastructure enabled secure authentication and proof of digital information authority for the platform. The plan's end goal is to set up a safe system that can detect and predict when social media posts may include FN.

In this research (Jiang *et al.*, 2021), assessed five ML models together with three DL models for performance over two datasets containing news items with varying length through hold out cross validation. Furthermore, they used embedding techniques to get text representation for ML models and TF-IDF and word frequency for DL models, respectively. To evaluate the models, they used F1-score, recall, accuracy, and precision. The researchers applied a McNemar modification to evaluate statistical differences within their results. They proceeded to use their novel stacking technique, which improved ISOT accuracy to 99.94% and KDnugget accuracy to 96.05%. In addition, when compared to baseline approaches, their suggested method performs quite well. Therefore, it is highly recommended for FND.

In this study (Nath *et al.*, 2021), tested and evaluated several ML and DL models for the purpose of detecting fake news. They made use of four databases. Their experiments revealed that RF and BoW achieved the highest accuracy of 98.8 percent on the FARN Dataset, whereas all other methods failed. Aside from that, they found that TF-IDF is the best feature extraction approach.

In this research (Ashraf *et al.*, 2021), reported a number of ML classifiers on the CLEF2021dataset for the n-gram-based news claim and topic categorisation tasks. On task 3a, they obtain an F1 score38.92% for classifying news claims, while on task 3b, they obtain an F1 score of 78.96% for classifying topics.

In this study (H. Ali *et al.*, 2021), tested four distinct deep learning architectures using the cutting-edge NLP attack library, Text-Attack, against a variety of adversarial attacks, including TextBugger, TextFooler, PWWS, and DeepWordBug. The architectures in question included MLP, CNNs, RNNs, and a newly suggested hybrid architecture, Hybrid CNN-RNN. They also investigate the effects on learnt model resilience of varying detector complexity, input sequence length, and training loss. According to their findings, RNNs outperform alternative designs in terms of robustness. In addition, they prove that the detector's resilience is enhanced when the input sequence length is increased.

(Priya and Kumar, 2021) said that a method for detecting COVID-19 FN using a deep ensemble is detailed in this paper. An ensemble classifier combines the strengths of a SVM, a CNN, and another classifier. Eight separate classical ML classifiers, each with its own set of n-gram TF-IDF words and character variables, are utilized to rigorously evaluate a proposed ensemble model. According to the research, n-gram features derived from characters outperformed n-gram features derived from words. Outperforming several other deep learning and standard machine learning classifiers, the proposed deep ensemble classifier achieved a weighted F1-score of 0.97.

(Khalil *et al.*, 2021) this is the first substantial collection of Arabic fake news, and it consists of 60,6912 pieces chosen from 134 Arabic internet news sites. A fact-checking tool in Arabic is used to categorise news sources as either not trustworthy, credible, or uncertain. Various ML approaches are employed for the detecting task. Deep learning models outperform their more conventional ML counterparts in experimental settings.

The entropy and complexity of the corpus were hinted at by the discovery of underfitting and overfitting difficulties during model training.

(Kousika *et al.*, 2021) emphasised their want to investigate if DL may succeed in identifying fraudulent online articles using content analysis alone. To achieve this objective, three separate neural network architectures are offered; one of these is built on Google's state-of-the-art language model, BERT. 'False news,' or misleading news items from reputable sources, may be detected with the use of NLP tools, which are the focus of this endeavour. They have built and assessed this method as a software system. Is it feasible to develop an algorithm that can distinguish among "false" and "genuine" news? Using this novel approach, SVM was able to detect FN with 92% accuracy, whereas NaiveBayes only managed 73%. This research found that using a novel method of prediction, the SVM classifier model performed better than the NBC model.

In this research (Mouratidis, Nikiforos and Kermanidis, 2021), presented a novel method for automatically detecting FN on Twitter that uses (a) pairwise text input, (b) a novel architecture for DNNs that enables tunable input integration across different network layers, and (c) word embeddings, language functions, and network account characteristics are among the many input options. An elaborate experimental setup uses both the news content and the innovatively split news headers to conduct classification tests on tweets. In terms of fake news identification, their primary findings demonstrate excellent overall accuracy. With a smaller feature set and fewer text embeddings from tweets, the suggested DL architecture beats the state-of-the-art classifiers.

In this research (Islam *et al.*, 2021), presented a unique perspective by assessing the credibility of news stories using NLP techniques. In this article, they provide a novel approach to news claim veracity assessment that employs ML for classification, position identification, and author credibility validation. Last but not least, the suggested pipeline employs a number of ML methods, including DT, LR, RF, and SVMs. Kaggle was the source of the FN dataset utilised in this research. The SVM method achieved an F1-score of 94.15%, an acc-uracy of 93.15%, a pre-cision of 92.65%, and a re-call of 95.71% in the experiments. The SVM outperforms LR, the runner-up classifier, by a margin of 6.82%.

In this research (Hamid *et al.*, 2020), sought to identify individuals spreading false information by analysing tweets pertaining to the COVID-19 pandemic and 5G belief systems. Text-based and structure-based FN identification are the two subtasks that make up the task. The first goal was to come up with six distinct solutions using BoW and BERTembedding. Even though most of the tweets about COVID-19 are categorised as either "5G conspiracy" or "others," three of the approaches approach the challenge as a binary classification problem by distinguishing between these two groups. As for the ternary classification job, their BoW-based technique achieved an F1-score of.606% on the development set, while their BERT-based method achieved an F1-score of .566%. Both the BoW-based and BERT-based algorithms achieved an average F1-score of.666% and.693% on the binary classification task, correspondingly. However, they depend on GNNs that averaged a ROC of.95% on the development set for structure-based fake news identification.

In this study (Sadeghi, Bidgoly and Amirkhani, 2020), relevant and comparable news articles provided by credible news outlets are utilised as supplementary information to deduce the credibility of a certain news story. Additionally, they compile and make available the first dataset for inference-based fake news identification, FNID, in two variants: FNID-FakeNewsNet, which has two classes, and FNID-LIAR, which has six classes. DTe, NB, RF, LR, KNN, SVM, BiGRU, and BiLSTM are among a classical and deep ML models that are enhanced using the NLI approach. Different word embedding methods like Word2vec, GloVe, fastText, and BERT are also utilised. After running the tests on the FNID-FakeNewsNet dataset (85.58% accuracy) and the FNID-LIAR dataset (41.31% accuracy), the suggested strategy shows an absolute improvement of 10.44% and 13.19%, respectively.

In this research (Goldani, Momtazi and Safabakhsh, 2021), planned to employ capsule NN for the purpose of FND. They employ several embedding models for news pieces varying in length. For shorter news items, static word embeddings are applied, but longer or more extensive news statements necessitate non-static word embeddings that provide gradual up-training and updating throughout training. Additionally, several n-gram levels are utilised for feature extraction. They test the designs they suggest using ISOT and LIAR, two popular and current datasets in the area. Impressive results when compared to modern techniques are shown by the LIAR dataset's test set, validation set, and ISOT, where an increase of 1% and a 3.1% improvement, respectively, were recorded.

In this study (Shaikh and Patil, 2020), proposed an approach to address the issue of FN by using several categorisation algorithms. Under resource constraints, the identification of FN becomes increasingly challenging. Databases and supplementary materials remain difficult to obtain. Classification methods such as SVM, NaïveBayes, and PassiveAggressiveClassifier have been utilised in this model. Their model achieves an accuracy of 95.05% in its output when employing feature extraction approaches like TF-IDF and SVM as a classifier.

In this research (Smitha and Bharath, 2020), showed a method and model for detecting fake news in news articles using ML and NLP. The proposed study designs feature vector by employing several feature engineering techniques like count vector, TF-IDF, word embeddings among others. This is due to the premise that seven different ML classification methods can be employed in identifying between real and fake news. Accuracy, F1 Score, recall, and precision are some of the metrics utilised to evaluate the model's performance.

In this study (Gravanis *et al.*, 2019), presented a model based on ML algorithms and content-based characteristics for detecting FN. In order to get the best model, they test out several feature sets that have been suggested for word embeddings and deceit detection. The analysis considers both the execution speed of the ensemble ML techniques AdaBoost and Bagging while testing prominent ML classifiers for their performance gains. With a wide range of historical data sources, feature sets and ML classifiers have been thoroughly tested and reviewed. In addition, they provide the "UNBiased" (UNB) dataset, a novel text corpus that combines news articles from several sources and satisfies a number of criteria meant to prevent biassed classification results. They find that it is possible to achieve high accuracy in classifying fake news using ensemble algorithms, SVM and a larger set of linguistic features with word embeddings.

In this research (Ye-Chan Ahn, 2019), identified the issue of determining the truth or falsity of an input sentence and any associated sentences retrieved from it using the Fact Data Corpus, where all phrases are supposed to be factual. They build a pre-training model tailored to Korean language employing cutting-edge BERT for the different NLP tasks. It is using this methodology that the data set identified by Korean FN is fine-tuned. The test set that was created using the fine-tuned model has an AUROC score of 83.8%.

In this study (Bauskar *et al.*, 2019), worked on a novel ML model that combines social news with content-based components in an effort to detect "fake news" using NLP techniques. The proposed model has performed well, averaging 90.62 percent accuracy and 90.33 percent F1 Score on a standard dataset.

In this research (Reis *et al.*, 2019), ran an extensive and exploratory study, using a wide variety of characteristics to generate hundreds of thousands of models. Since these models' characteristics are selected at random from the available pool, they are impartial. Even while most models fail miserably, they did manage to create a few that successfully distinguish between real and fake news by producing extremely accurate choices. Models that assign a higher probability to a randomly selected fake news story compared to a randomly selected truth were the primary focus of their investigation.

In this study (Verma, Mittal and Dawn, 2019), suggested utilising GRUs, LSTM, and RNNs to test categorisation. The results showed promise. In order to see the neural network and put the suggested framework into action, Tensorboard is utilised. The classification results and confusion matrix suggest that the LSTM model has the potential to achieve an accuracy score of 94%. The time investment is more, but it's worth it. A high-quality training set is essential for establishing the fake news using the learning model.

In this research (Shu, Wang and Liu, 2019), recommended using social context to identify fake news, a fresh challenge. The tri-relationship embedding framework they proposed, TriFN, classifies fake news by simulating interactions among users and news, as well as among publishers and news. They show that the suggested strategy much beats previous baseline approaches for FND in studies conducted on two real-world datasets.

In this study (Shu, Wang and Liu, 2018), constructed databases that investigated a reliability of FN and separated people into two categories: "experienced" users, who are adept at identifying false information, and "naïve" users, who are more likely to believe it. Their research shows that by comparing explicit and implicit profile attributes, these user categories may help identify fake news. This paper's results provide the groundwork for studies on automated fake news identification in the future.

In this research (Roy *et al.*, 2018), purpose of identifying FN and sorting it into pre-established fine-grained categories, many DL models have been proposed to be developed. At the outset, they build models using CNN and Bi-LSTM connections. For the final classification, a MLP is fed the representations derived from these two models. Their trials on a benchmark dataset demonstrate encouraging outcomes, surpassing the state-of-the-art with an overall accuracy of 44.87%.

In this research (Al Asaad and Erascu, 2018), proposed a specific kind of supervised learning for identifying fake news. They trained a ML model using a dataset that included both actual and fraudulent news using the Scikit-learn package in Python. Text representation methods including BoW, TF-IDF, and Bi-gram frequency were utilized to extract the features from the dataset. They identified clickbait and authentic/fake material using probabilistic and linear classification, respectively. Their research showed that when using the TF-IDF model for content categorization, linear classification yields the best results.

Fake News in Social Media Content Analysis Using Machine learning

In this research (Ayenew *et al.*, 2024) used DL to detect Amharic FN by combining genuine and fake articles with the Amharic dataset. They classified fake and authentic news in Amharic using GRU networks and LSTM. The GRU model maintained a 98% success rate in detecting fake news which proved to be the most effective. The vanishing gradient problem, a frequent issue with RNNs, is well handled by GRUs.

In this study (Maham *et al.*, 2024), to build a proposed end-to-end system for FN detection strength of adversarial training is employed. ANN: Adversarial News Net functions as a system that utilizes the acronym to extract emoticons from datasets to understand their meaning regarding fake news. An identification capability of FN by a model improves through this process. The researchers applied ANN framework testing to four open-source datasets and proved its superior performance compared to baseline methods and the findings of previous studies through adversarial training. Results from the trials that included Adversarial Training showed a 2.1% performance boost over the RandomForest baseline and a 2.4% performance boost over the BERT baseline method.

In this research (Garg, Saudagar and Gupta, 2024), proposed an approach for identifying FN on social media. This approach uses N grammes and Word2Vec to extract TF-IDF features in two ways: (i) using the stemming technique and (ii) without. In order to identify fake information, both of these processes are carried out and fed into supervised ML algorithms like LR, RF, SVM, GB, Adaboost, and SGD. Results from using random forest on the unigram outperform those of the bigram and trigram, according to the evaluation. Trigram surpassed all other categorisation methods. Using or not using a stemming approach, Unigram is always more precise. at the provided dataset, Word2vec's accuracy at detecting fake information is lower.

In this study (Madhunitha, 2024), performed extensive research into the efficacy incorporates a wide range of misleading material in visual, textual, and audio forms. They then decide their dataset meticulously and compare them with four classifiers including SVM, RF, MLP, and NB. By combining the predictions from these many models, they offer a voting classifier that is more accurate and consistent. Peculiarly, the generalization capability of the Voting Classifier emerged when it revealed the robustness of the ensemble method and reached the perfect accuracy. This work makes an attempt to demonstrate how ensemble methods can be used to build a strong foundation for identifying fake news systems and provide a defense against the continuous propagation of disinformation.

In this study (Rampurkar and Thirupurasundari, 2024), seeks to enhance digital content fake news detection capabilities for better information reliability along with integrity in digital environments. The investigators begin by compiling a large database of news stories from both legitimate and fake sources. These methods include stemming, tokenization and stop word elimination which prepare raw text for feature extraction purposes. To determine how important each word is in each article, the feature selection phase use the TF-IDF algorithm. Classifying news articles as real or fake using the Naive Bayes algorithm is the next stage. The Naive Bayes approach uses a probabilistic method that assumes the words' attributes are conditionally independent to calculate the article's category likelihood. A collection of pertinent textual elements are used using Logistic Regression to estimate the likelihood of a news story being fake or real. A primary goal of logistic regression is to develop effective categorization of news items between real and fake through feature engineering and model evaluation methods. The efficacy of the related methods is established by evaluating the model's correctness using the confusion matrix.

In this research (Dhanuka and Tiwari, 2024), proposed developing a product model based on supervised ML algorithms to authenticate fake news through python libraries such as scikit-learn and NLP tools alongside relevant research. They recommend using the scikit-learn package in Python for feature extraction and other useful tasks, like as working with tiff and count vectorizers. Then, they will choose the best-fit features through experimentation, utilising the confusion matrix findings as a guide, in order to get the maximum level of precision.

In this study (Al-alshaqi and Liu, 2024), presented a complete methodology which uses ML and DL techniques to detect FN across text documents alongside images and video contents. Initially, the study uses a number of classifiers to examine textual data. The approach adds a variant of CNN to process visual data and BERT to evaluate text information through a hybrid system. The effectiveness of the suggested models is shown via trials using the MediaEval 2016 corpus for photo verification and the ISOT dataset for fake news. The RFC surpassed competing algorithms with a 99% success rate when applied to textual data. Compared to baseline models, the multimodal approach performed better, and it improved accuracy by 3.1% compared to previous multimodal methods.

In the research (Jouhar *et al.*, 2024), used evaluative variables like accuracy and precision to present a holistic perspective. This strategic research determines which machine learning technique is more effective in classifying articles as either fake or real news. A literature review is only one of many topics covered; others include training models, selecting metrics, optimising and evaluating models, vectorisation, dataset selection, data preparation and cleaning, and more. Beyond technological competence, it has a positive effect on media trust, democratic procedures, and the veracity of material. Many groups stand to gain from an all-encompassing strategy, including media outlets, social media, and public agencies. Significantly improves information believability through thorough review.

In this study (Nidha *et al.*, 2024), noticed that browsing news feeds is a common activity of social media. A more serious issue, though, is the veracity of the news stories that people are seeing and sharing online. The alarming spread of pseudo-news and hoaxes is reminiscent of an uncontrollable the pandemic. People used to share news that seemed satisfactory without verifying the source's credibility or determining whether it was real or fake. They were unaware that these careless actions could weaken the influence of fake news and cause serious problems, particularly during disasters. Google is one of a group of reputable sites that they maintain. They will get useful information when they seek for authentic material on these sites. They will get results that are irrelevant and mismatched if they look for fake stuff online. This program finds the desired news story by searching the web and compiling results from various web pages or Google links. Afterwards, they match the search results with the relevant news articles. Understanding items and comparing them to identify commonalities is not a simple process. That is accomplished by use of methods such as NLP and related algorithm sets. if or not there are parallels indicates if the news is authentic or fake.

In this research (Salh and Nabi, 2023), three different ML and DL classifiers (RF, SVM, and CNN) were utilised to detect the fake news dataset, and three suggested techniques—word embedding, TF-TDF, and count vector—were put into practice to extract features from news texts. The research findings indicate that CNNs demonstrate a high level of effectiveness when detecting FN based on text content. Compared to the other models tested in the research, CNN performed better, with an F1-score of 95% and an acc-uracy of more than 91%. In lighter drama, this paper demonstrates that even in adverse conditions, it is possible for ML algorithms to detect FN in LRIs, including Kurdish.

In this research (Hisham, Hasan and Hussain, 2023), concentrated on informing the targeted audiences about the problem of FN on socialmedia and sharing an advice on how to distinguish the sources of reliable information. It delves into ways to identify online social network sources, writers, and topics of FN. After establish a veracity of news stories, the project makes use of an open-source internet dataset that contains both actual and fake news. The article discusses different methods for extracting text features and algorithms for classification. A most effective method, which achieved an accuracy99.36%, was the SVM linear classification algorithm that made use of TFIDF feature extraction. The accuracy scores for RF and NB were 94.74% and 98.25%, respectively.

In this research (Mohawesh *et al.*, 2023), stated that, despite a rise in multilingual internet content, low-resource languages still have difficulties in detecting fake news since there aren't enough annotated corpora and approaches. To get over these problems and solve the issue of multilingual fake news detection, they come up with a new representation learning framework based on semantic graph attention to extract structural and semantic properties of texts. In tests conducted utilising TALLIP fake news datasets, their suggested approach outperformed state-of-the-art algorithms for the multilingual FND problem, with accuracy increases ranging from 1% to 7%.

In this study Singh & Selva (2023), investigated several ML classifiers for a purpose of identifying FN. A piece of news may be identified as authentic or fake based on linguistic features of the news dataset that was taken from Kaggle. With these features in place, they may test how well the acquired dataset performs during model training using various ML classification techniques.

In this research (Singh and Selva, 2023), proposed Hindi news pieces culled from a variety of news outlets. Every step of the process, from preprocessing to feature extraction to classification and prediction, is covered thoroughly. The fake news is detected using several ML techniques, including Naïve Bayes, LR, and LSTM. In the preparation phase data gets cleaned and stop words are removed while tokens get split into individual units then subject to stemming. Feature extraction is best accomplished using the TF-IDF approach. They examine the use of NaïveBayes, LR, and LSTM classifiers for identifying FN, as well as the probability of truth. The LSTM classifier outperformed the other two with a 92.36 percent accuracy rate.

In this study (Adedoyin and Mariyappan, 2022), goal of this research is to examine, compare, and evaluate the effectiveness of various DL and ML algorithms in identifying and rejecting FN reports. This study aims to create and test seven models, including two DLmodels and five MLmodels, to determine which is the best at spotting fake news. One of these metrics, ROC curve, acc-uracy, re-call, and pre-cision are calculated to see how good the model is. "FakeNewsDetector" is one web application that helps the users to identify whether the news is fake or not.

In this research (Khanam *et al.*, 2021), presented a method which employs ML techniques to detect FN. Postings and sequence segments with TF-IDF values acted as features to extract information which was classified using SVM methodology. The proposed testing method with real and fake news data supported the researchers' findings to prove their methodology. The obtained results confirm the system's effective performance.

In this study (Mahmud *et al.*, 2022), used a few well-known ML techniques with graph neural networks to determine how fake news spreads on social media. Utilising just text data, they apply many pre-existing machine learning methods to the UPFD dataset. Additionally, they construct several GNN layers to integrate graph-structured data on news propagation with text data used as a node feature in them. In their research, GNNs provide the best solutions to the challenge of detecting FN.

In this research (Ahmed, Hinkelmann and Corradini, 2022), employed three classifiers: PassiveAggressive, NaiveBayes, and SupportVectorMachine, to identify FN using ML techniques. Classification algorithms aren't tailored to fake news, thus simple classification isn't always the best approach for detecting them. By fusing ML with text-based processing, they are able to identify FN and develop classifiers for news data. Text classification primarily aims to discover various text characteristics and then use those features for categorisation. There is currently no viable method to distinguish between fake and non-fake material because to the absence of corpora, which is a big difficulty in this area.

In this study (Senhadji and Ahmed, 2022), focused on seeking to ascertain the truthfulness of news articles via an employ of ML andDL methods. The objective of this study is to positively detect FN employing NB and LSTM classifiers. The results show that compared to naïve bayes, LSTM obtains a 92% accuracy rate. When compared to the outcomes of the linked work, the proposed approach's findings are significantly better.

In this research (Alghamdi, Lin and Luo, 2022), investigated in order to deduce the causes of and solutions to a problem of FN. In particular, they trained a benchmark using a varietyof (1) classical MLalgorithms like LR, SVM, DT, NaiveBayes (NB), RF, and XGB, as well as an ensemble learning method combining these algorithms, (2) advanced MLalgorithms like CNNs, BiLSTM, BiGRU, CNN-BiLSTM, CNN-BiGRU, and a hybrid approach combining these techniques, and (3) DLtransformer-based models like BERT$base$base and RoBERTa$base$base. This study evaluates several methods on four popular real-world FN datasets: LIAR, PolitiFact, GossipCop, andCOVID-19, using pretrained word embedding algorithms. They also compare context-independent embedding approaches (like GloVe) to BERT base—contextualized representations—to see how well it identifies fake news.

In this study (Hatwar *et al.*, 2022), attempts to compile news stories and then determines whether they are authentic or fake using ML models and NLP techniques. For the purpose of identifying "fake news," Sor deceptive news reports that originate from unidentified sources, three NLP systems have been developed. Using a TF-IDF matrix or a count vectorizer as the only building blocks for your model can only take you so far. However, crucial features like word order and context are ignored by these models. Because the Python sci-kit-learn package provides helpful utilities like CountVectorizer, Tf-IDF Vectorizer, and HashingVectorizer, they proposed utilising this library to extract features from text data.

In this research (Jadhav *et al.*, 2022), "Blacklists" of questionable sources and writers have been proposed as the most often used of these efforts. In order to give a more complete end-to-end solution, these methods are useful, but they must take into account more complicated circumstances when reputable sources and writers create fake news. Finding a way to differentiate among real and fake news via the analysis of language patterns utilising ML and NLP was the driving force behind this research. The project's outcomes prove that ML can be effective for this specific task. Together, the model and the application help visualise the categorisation decision, and they pick up on numerous intuitive signs of true and fake news.

In this study (S. Khan *et al.*, 2022), proposed using a dataset for categorisation purposes that is a combination of COVID-19-related news stories pulled from a variety of social media and news sources. To begin, the dataset undergoes preprocessing to eliminate any extraneous text. Then, the raw text data is subjected to tokenisation in order to extract the tokens. Feature selection is then carried out to prevent the computational burden of analysing every feature in the dataset. Afterwards, the characteristics related to language and emotion are retrieved. Lastly, advanced MLalgorithms are taught to categorise the dataset associated with COVID-19. After then, a number of measures are used to assess these algorithms. A higher accuracy rate of 88.50% was achieved by the RFC, which is superior than the other classifiers.

in this research (Kushwaha and Singh, 2022), proposed an employ of ML techniques to uncover inaccurate information. they evaluate and contrast three distinct phases of ML approaches. In addition, they will be utilising three distinct models: RandomForest Classification, Decision Tree Classifier, and LogisticRegression. Their project's findings indicate that they have successfully acquired your several levels of accuracy in a sequential manner. Finding out if the provided information is real or fake might be very beneficial to their endeavour.

In this research (Pranay Patil, Abrar Khan, 2021), fake news can cause confusion and mislead people. Based on this, they create a FN detector in Python with the help of packages like pandas, sklearn, and matplotlib. ML algorithm is being used to identify instances of false news in relation to the US elections, and it will then use this data to determine if a piece of news is fake or not. The logistic regression model they're utilising has a 98% success rate.

In this study (Anjali Gangan, Sayali Wagh, Wajida Siddiqui, 2021), "Blacklists" of untrustworthy sources and manufacturers are one of the most popular of these endeavours. Complete arrangements may be made from start to end using these tools, but they must address the more difficult circumstances where more reliable sources and authors spread information about counterfeiting. Using AI, regular language preparation techniques, and other methods, this company aimed to create a device that could detect language maps that represented authentic and fake information. It is clear from this project's outcomes that there is a considerable ceiling for machine learning and AI. In addition to developing a model that can detect many natural indicators of fake and real news, they also created an app to help with the representation of the categorisation option.The technique for recognising fake news involves simultaneous interactions between users and news publishers.

In this research V et al. (2021)intended to gain insight into the reasons underlying the accidental spread of perhaps false material with the hope of assisting in the detection and suppression of fake news.The only thing that has gotten you this far is a model that supports a count vectorizer or a TF-IDF matrix. These subsequent models, however, occasionally failed to take into account crucial elements like word order and context. Despite having comparable word counts, two articles may have very different meanings.

In this study (Hegde and Shashirekha, 2021), proposed the MUCS team, which describes the ensemble of four ML classifiers—RandomForest (RF), MLP, GradientBoosting (GB), and Adaptive Boosting (AB)—submitted to UrduFake 2021. Training the ensembled classifier using word uni-grams, character n-grams, and fastText Urdu word vectors yielded a suggested model that placed twelfth in the shared task, with an accuracy of 0.713 and a macro F1-score of 0.552.

In this research (Gupta and Meel, 2021), proposed Passive-Aggressive Classifier is put into action. This technique was used to two databases: one holding real news and the other having fake news. This article explains how to use an ML system to combat the issue of fake news. To detect fake news, a variety of classifiers are used. The Passive-Aggressive Classifier reaches an impressive 97.5% accuracy rate, as shown by the experimental data.

In this study (Khanam *et al.*, 2021) create a supervised ML algorithm-based product model that can check the veracity of fake news utilising resources like scikit-learn, NLP for text analysis, and studies on detecting such stories. The Python scikit-learn package, which has useful methods like CountVectorizer and TiffVectorizer, is suggested for text data tokenisation and feature extraction. Extracting features and vectorisation are the end results of this strategy. next the measurement of fit and precision using the confusion matrices, the next step is to identify the most effective features using feature selection methods.

in this research (Gundapu and Mamidi, 2021), recommended documenting an approach to assess the credibility of COVID-19-related social media posts. Our most effective method for identifying fake news relies on a combination of three transformer models (BERT, ALBERT, and XLNET). The "COVID19 FakeNewsDetection in English" shared challenge from ConstraintAI 2021 used as the basis for training and evaluation of this model. Out of 160 teams, their system ranked 5th with a f1-score of 0.9855 on testset.

In this study (Bangyal *et al.*, 2021), proposed a very accurate method for COVID-19 fake news detection. They began with data preparation (replace the missing value, noise reduction, tokenisation, and stemming) because the fake news dataset includes fake news about COVID-19. They used an IDF weighting and word frequency weighting semantic model to describe the data. They used four DL methods (CNN, LSTM, RNN, and GRU) and eight ML algorithms (NB, Adaboost, KNN, RF, LR, DT, neural networks, and SVM) in the measurement and evaluation stage. After that, they used the results to build a powerful Python prediction model. The classification model was trained and evaluated using performance indicators such as classification rate, confusion matrix, and true positives rate. Last but not least, they put the algorithm through its paces by determining the sentiment class of each unclassified COVID-19 fake news story.

In this research (Reddy *et al.*, 2020), took a text-only approach to the challenge of identifying fake news, ignoring all other pertinent information. In their study, they found that using ensemble approaches in conjunction with stylometric characteristics and text-based word vector representations could achieve a prediction accuracy of 95.49% for fake news.

In this study (Ahmad *et al.*, 2020), proposed a method for automatically categorising news items using an ensemble approach to ML. After differentiate between true and fake information, they investigate several textual features. Based on these features, they use an ensemble setup to train multiple ML algorithms, which are then tested on four real-world datasets to determine their performance. Their proposed ensemble learner method outperforms individual learners, according on experimental assessment.

In this research (Abel Nyambe Mushiba, 2020), was advised to develop and deploy an anti-fake-news solution for use in supervised ML product model construction. This work will mainly use an NBC to construct a model that can distinguish between authentic and fake news based on the words and phrases used in it. Tools like a count vectorizer (using word tallies) or a TF-IDF matrix will be utilised. It is quite probable that two papers with comparable word counts will have entirely distinct meanings.

In this study (Yuffon *et al.*, 2020), proposed a FND algorithm that addresses the issue by utilising a TextVectorizer and ML approaches. The most effective feature extraction method and classifier in the experimental assessment was TF-IDF, which achieved an accuracy of above 97%.

In this research (Liu and Wu, 2020), to quickly identify fake news, a revolutionary deep neural network was proposed. It comprises three new elements: (1) an extractor for status-sensitive crowd responses that takes into account both the text responses of users and their associated profiles to derive characteristics about those individuals, (2) an attention mechanism that is aware of its position and prioritises user replies based on their ranking, and (3) a method for doing feature aggregation using several window widths using multi-region mean-pooling. The researchers achieved superior results using their model against standard reference points by reaching a 90% success rate to detect false news during its introductory 5-minute period before 50 retweets. This is a substantial improvement. The method operates efficiently under PU-Learning requirements using 10% fake news examples to provide effective results.

In this study (Jadhav and Thepade, 2019), put forward a system that employs the use of DeepStructuredSemanticModel and improved RNNs for the detection of articles of fake news. Accuracy of 99% was achieved by the proposed method, which intuitively recognises crucial traits associated with fake news without prior domain expertise. Accuracy, specificity, and sensitivity form the basis of the performance analysis approach utilised for the proposed system.

In this study (Tschiatschek *et al.*, 2018), decentralization mechanism successfully prevented fake news from spreading through the network which lowered the distribution of false information. This goal is particularly difficult to accomplish since it calls for the rapid and certain detection of fake news. They demonstrate the need of understanding users' flagging accuracy for effective usage of user flags. Users learn about accuracy levels with Bayesian inference methods that detect fake news through the innovative system called Detective. Their system makes advantage of posterior sampling to actively balance two goals: exploitation, which is choosing news that maximizes objective value at a specific epoch, and exploration, which is choosing news that maximizes information value for learning about users' flagging accuracy. They highlight the power of using community signals for fake news identification by conducting comprehensive tests to prove the success of their method.

In this research (Aldwairi and Alwahedi, 2018) This project aims to provide a mechanism that consumers may use to identify and avoid websites that give incorrect or fraudulent information. For the purpose of properly identifying fake postings, they employ basic and meticulously chosen characteristics of the header and content. Using the logistic classifier, the experimental findings demonstrate an accuracy of 99.4 percent.

2.3 Critical Evaluation of Past Methodologies

Many different NLP approaches have been used in the past for the purpose of detecting fake news, each with its own set of benefits and drawbacks. Fast training durations, interpretability, and simplicity are the hallmarks of traditional machine learning methods including SVMs, LR, DT, and TF-IDF or BoW features. However, these models struggle with capturing semantic relationships and contextual meaning, making them less effective in detecting nuanced or deceptive language. Deep learning methods, including CNNs, LSTMs, and hybrid architectures, have improved performance by learning complex patterns from data. These models excel at handling sequential and hierarchical data structures but require large annotated datasets and are computationally expensive. Transformer-based models like BERT, Roberta, and XLNet have set new performance benchmarks due to their contextual understanding and transfer learning capabilities. Nonetheless, their dependency on pre-trained language models trained on generic corpora can introduce bias, and their deployment at scale poses challenges in terms of latency and interpretability. Zero-shot and few-shot learning models, such as ZS-FND, offer promising results in scenarios with limited domain-specific data but require further validation across languages and platforms to assess their reliability and robustness.

2.4 Behavioral and Media Theories in Fake News Spread

To fully understand the spread of fake news, it is essential to incorporate behavioral and media theories that explain how and why individuals consume and share misinformation. The **Elaboration Likelihood Model (ELM)** provides insight into how people process information either through a central route (deep processing) or a peripheral route (superficial cues), suggesting that individuals may accept fake news without critical analysis when cognitive effort is low. Similarly, **Selective Exposure Theory** posits that users tend to engage with content that aligns with their pre-existing beliefs, reinforcing echo chambers that make the spread of false information more potent. The **Uses and Gratifications Theory** further explains how individuals actively seek content that satisfies their informational, emotional, or social needs—sometimes at the cost of accuracy. Integrating these theories into fake news detection frameworks can help in designing more context-aware models that not only focus on textual patterns but also incorporate user behavior, credibility signals, and dissemination patterns to better capture the social dynamics of misinformation.

2.5 Dataset Representativeness and Bias

The datasets used to train and evaluate models for fake news detection have a significant problem with their representativeness and possible bias. Many commonly used datasets, such as Buzzfeed, ISOT, and Fakeddit, are either domain-specific or linguistically limited, raising concerns about generalizability across different cultures, languages, and social contexts. Several studies attempt to address these issues by augmenting datasets (e.g., PolitiTweet) with region-specific linguistic and cultural features or by combining multiple sources to improve diversity. However, problems such as class imbalance, outdated content, and curated sampling remain. These biases can cause models to overfit to specific styles or topics of fake news, reducing their effectiveness in real-world scenarios where misinformation is constantly evolving. To enhance robustness and fairness, future work should focus on building large-scale, multilingual, and temporally updated datasets with better class balance and diverse content sources, while also incorporating metadata and user interactions that reflect real-world usage.

2.6 Research Gap

A review of the literature on social media FN identification finds some interesting and potentially useful applications of ML and DL. Innovative techniques that come with the transformer versions, a combination of both transformer and recurrent networks, and better embeddings show improvements in accuracy. However, there still emerge some deficiencies in establishing general solutions that can be easily adjusted to such changes because of differences in languages and contexts of various territories. The previous works can learn performance-oriented improvement for unique datasets but they are not efficient while addressing the diverse and dynamic features of fake news. Furthermore, there is a lack of investment in developing multi-modal FND systems that use social network characteristics in addition to textual and visual information. These shortcomings need to be remedied in order to create global systems that are both methodical and resilient enough to combat fake news.

CHAPTER III:

RESEARCH METHODOLOGY

3.1 Proposed Methodology

Today's society is negatively impacted by the dissemination of fake and misleading information on blogs and social media platforms. The news is skewed with false information and enhanced with questionable facts, leading to harmful social panic and interpersonal anxiety(Prachi *et al.*, 2022). Misinformation like this undermines public trust in news outlets and has far-reaching consequences for crucial political processes like elections and the stock market. To manually detect the dissemination of fake and edited news, human verification is often performed. This method of manually verifying facts is inefficient, time-consuming, difficult, and prone to subjectivity. Automatic methods that use deep learning and NLP algorithms have been utilized to detect fake news in recent years(Veeraiah *et al.*, 2024). Thanks to advancements in technology and AI, these automated solutions efficiently prevent the spread of disinformation and fake news. Future advancements in the detection of fake news using these methods have piqued the curiosity of researchers.

The proposed methodology for detection and classification of fake news on social media includes a sequence of simple and well-ordered steps, which consist of the following steps: Initially, the datasets PolitiFact and Gossip Cop have been collected from open-source GitHub. After the data is collected, the data passes through a rigorous preprocessing process, starting with assigning the key attribute of fake or real news to each dataset. The datasets are then concatenated and cleaned, and null values are removed in order to maintain data quality. Text preprocessing comes next and involves elements such as converting all titles to lowercase, handling of punctuation and elimination of URLs. Tokenization is done for the purpose of dividing the text into segments retrievable for additional processing, elimination of titles that are recurrent to avoid repetition. To control the class imbalance, random oversampling data balancing technique is used. EDA is performed to reveal hidden patterns and insights. The data is split into an 80:20 ratio for training and testing the model to ensure its accuracy. A sequence classifier model with Roberta architecture is used as a base model, with an optimized choice of hyperparameters. Related to the assessment of the model's performance, the measures include accuracy, precision, recall, F1score, and their formatted representation shown in confusion matrix, AUC. The last thing to do when building a model to identify fake news is to put it to use by labelling news pieces as either real or fake. Figure 3.1 shows the proposed system's flowchart for detecting fake news.

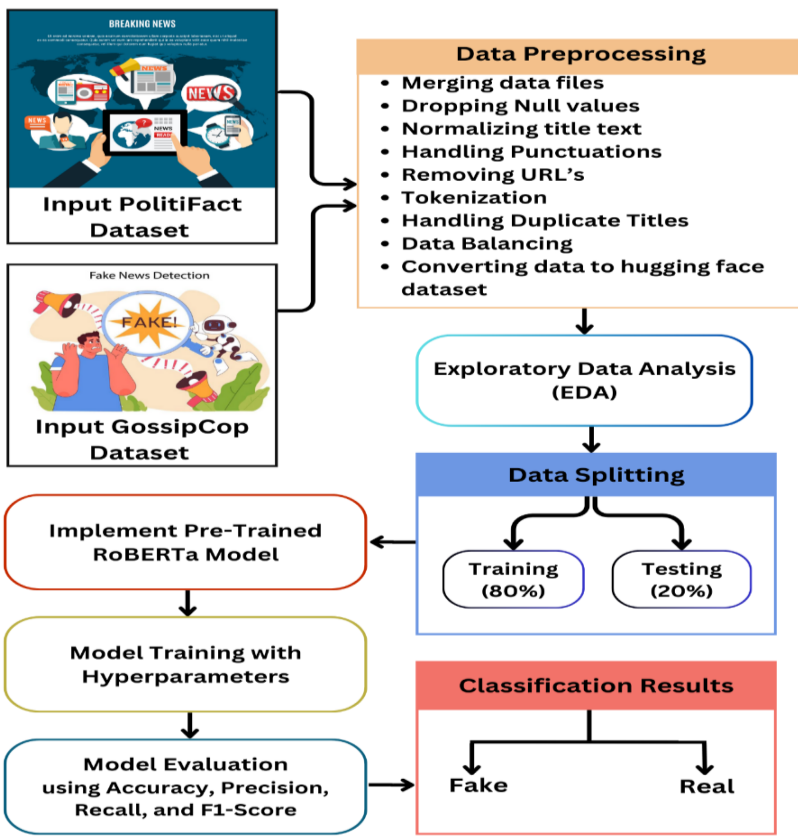

Figure 3.1: Flowchart for Proposed Fake News Detection System

Data Collection

Data collection is an important and initial phase in machine learning task(Roh, Heo and Whang, 2021). This research used the Fake News Net dataset, which is available as open-source software on GitHub, to identify and categorize social media fake news. The dataset contains two separate datasets, PolitiFact and Gossip Cop. It has information about the fake and real news. The collection includes graphical and textual information, all tweets and retweets for each news item, and user details for the relevant Twitter accounts for each item. The specific description of these datasets are given below.

PolitiFact Dataset

The PolitiFact dataset has a part of Fake Newsnet dataset. It contains the total rows of 762 and 5 columns. The dataset supports three key functions including fake news recognition, Twitter content distribution investigation and public news involvement research through article data and social media interaction relations.

Gossip Cop Dataset

The Gossip Cop dataset has also a part of Fake Newsnet dataset. It contains the total rows of 20465 and 5 columns. The combined data collection enables research analysis at three levels which includes article information and social media reactions and enables false news detection along with social media content research and user participation evaluation.

Justification of Dataset Selection and Limitations: While PolitiFact and Gossip Cop datasets are widely used and sourced from the Fake Newsnet repository, their scope is limited to English-language content and primarily focuses on political and entertainment news, respectively. This selection may introduce dataset bias, limiting generalisability to other news domains (e.g., health misinformation or international news). These datasets also lack multimodal content, such as images and videos, which are common in social media posts. The inclusion of richer multimodal datasets (e.g., Weibo, Twitter-15/16, or faked it) could potentially improve the detection of fake news spread through memes, manipulated media, or viral videos.

Data Pre-Processing

The data preparation approach consists of various techniques that enhance original data quality by eliminating outliers and completing missing value gaps(Fan *et al.*, 2021). In this study, performed various preprocessing steps including labeling data, merging, shuffling, cleaning, duplicate handling, text embedding, and data balancing on both the PolitiFact and Gossip Cop datasets, to make the data efficient for further classification and detection of fake news. These preprocessing steps are given below:

- **Labeling Datasets:** Data labelling stands as the essential initial step which supervised ML requires to create and evaluate its prediction models. Applying dataset classification techniques to the PolitiFact and Gossip Cop fact-checking services, the research sought to identify instances of fake news. The dataset for "PolitiFact Fake" contained false news and was identified by value "0" but "PolitiFact Real" had value "1." In the "Gossip Cop Real" dataset the value 1 indicated genuine news content along with the "Gossip Cop Fake" dataset that received the value 0 for news contents deemed fraudulent. This labelling approach through a binary "label" column added to all datasets allowed proper data integration and representation.
- **Merging & Shuffling Data Files:** To ensure a consistent structure and reduce sequential bias while training models, it is necessary to merge and shuffle data before putting them into a machine learning environment. In this research, a Data Frame named GC_Combined was created by merging the "Gossip Cop Fake" (GCF_Data) and "Gossip Cop Real" (GCR_Data) datasets by data concatenation.

In order to keep things consistent and prevent index misalignment, the indices of the combined Data Frame were reset. The GC Combined Data Frame was randomly shuffled with the index set to reflect the scrambled state to further increase the data's quality and avoid any order-dependent bias. A properly integrated and impartial dataset, prepared for efficient training and assessment, is the result of these preparation procedures.

- **Data Cleaning:** The essential step before analyzing a dataset requires data cleaning operations to eliminate duplicate and unnecessary records. To ensure the data remained intact, this research detected rows in the combined dataset that included missing values and removed them. In order to keep the text representation consistent, the title column underwent text normalization, which included changing all characters to lowercase. In order to eliminate URLs and punctuation marks, which might potentially interfere with natural language processing, custom routines were used. Lastly, the title column underwent tokenization to reduce the text to smaller units, such words or subworlds, which allowed for subsequent processing processes to be more efficient.
- **Handling Duplicate Titles:** Duplicated datasets can result in duplication of data and model over fitment, which decrease the accuracy of utilizing the machine learning algorithm. This research detected and eliminated duplicate items based on the title column to guarantee the dataset included only unique news titles. The training data obtains diversity and impartiality through this measure which leads to improved model performance.
- **Text Embedding:** In order to convert the text data into a format that is appropriate for machine learning models, BERT embeddings were implemented in the title column. This process required the creation of a custom function

get_bert_embedding, which converts the titular string into a vector of fixed size. These embeddings captured meanings and context of the text thus ensured the generation of a highly meaningful numerical representation for each of the news title which is crucial in subsequent steps.

- **Data Balancing:** Data balancing remains important because it protects models from developing class bias for majority classes in binary classification systems (Mujahid *et al.*, 2024). In this study, the class imbalance was addressed by resampling the smaller category (either fake or real news) to match the size of the larger category. Balancing the dataset through this method allowed the model to acquire equal knowledge from both classes thereby enhancing its predictive capabilities on new data.

Exploratory Data Analysis

The main component of EDA is data visualization while interactive display and exploration along with data discovery of trends and behaviors and correlations also constitute its core elements (Semanjski, 2023). The purpose of the presentation, which is an early stage of exploratory data analysis, is to provide the audience a quick overview of the dataset. Descriptive statistics are computed and visually represented using a variety of visualization techniques (e.g., histograms, scatter plots, bubble charts, matrix plots, etc.) according to the kind of data in the variable.

Train-Test Split

A popular method for validating models is data splitting, in which researchers divide a given dataset into two distinct sets: training and testing (Joseph, 2022). For the present study, the datasets were divided into training and testing based on 80:20 proportions where 80% was used for training while 20% was used for testing. This approach guarantees the model dataset diversity, in the sense that there is enough data for learning; and, at the same time, there is even another set for testing its performance without influence.

Deep Learning Techniques

For the identification of fake news, deep learning algorithms like CNNs and RNs have recently been deployed (Kolev, Weiss and Spanakis, 2022). These algorithms can better understand article context and learn complicated representations of the articles. Nevertheless, these techniques need copious amounts of data for efficient training and can be somewhat costly computationally.

Roberta Model

Several methods have enhanced BERT's power, despite its already impressive performance on a variety of NLP tasks. An example of such a piece of work is the Robustly Optimized BERT Pre-training Approach, or Roberta. Researchers from Facebook and Washington University proposed this variant of BERT. Roberta excels in both optimizing the pre-training of BERT architecture and predicting purposely obscured areas of text. Aside from certain tweaks to the training process, Roberta employs the identical architecture as BERT. It modifies critical BERT hyperparameters(Kitanovski, Toshevska and Mirceva, 2023). Instead of using BERT's next-sentence pre-training purpose, it just employs MLM. A number of factors contribute to Roberta's superior performance and performance-oriented approach to masked language modelling compared to BERT. These include a larger vocabulary (approximately 50,000 words), faster learning rates, longer training sequences (512 tokens), and dynamically modifying the masking pattern, as opposed to BERT's static mask(Angin *et al.*, 2022). The following section provides the preprocessing and training of the proposed Roberta model.

Model Preprocessing

This research used a pre-trained Roberta model (roberta-base) sourced from Hugging Face's model hub for the sequence classification task. The tokenizer was loaded with the model to transform raw text into tokens suitable for the model. A preprocessing function was established to tokenize and pad the text data into a uniform length, assuring alignment with the model's input specifications. The training and testing datasets were transformed into Hugging Face's Dataset format, with the preprocessing function implemented on both datasets.

Model Training

In order to maximize performance, the model training was set up using certain parameters. Model checkpoints and tensor board files will be saved to the. /results directory. At the conclusion of each epoch, the evaluation and saving techniques were configured to take place, guaranteeing that checkpoints are preserved and that the optimal model is chosen according to test loss. With fp16=True, we enabled 16-bit floating-point precision to speed up training. Both the training and assessment batches had a size of 16, and the learning rate was set at 2e-5. For regularization, the model was trained for 15 epochs with a weight decay of 0.01. Logs were captured every 50 steps and saved in the. /logs directory for training purposes. Automatic model loading was done according to the F1 score, and no more than two checkpoints could be stored.

Model Evaluation

The efficiency of developed model was measured by checking on the following aspects: accuracy, precision, recall, and F1 score. A confusion matrix was used in order to observe the classification outcome and the ROC curve was used for evaluating the model’s capacity to classify between classes. Last, the training loss and the evaluation metric of the model were taken to give an overview of the performance of the model and how the model generalizes on new data.

Consideration of Alternative Architectures

While Roberta is selected for its proven performance in language understanding tasks and its superior results in benchmark datasets compared to BERT and XLNet, alternative architectures like Graph Neural Networks (GNNs) could offer advantages. GNNs can capture relationships and user interactions in social networks, enabling better context-aware predictions. Additionally, hybrid models combining Transformer architectures (e.g., Roberta) with sequential models like RNNs may enhance temporal analysis of news propagation and offer deeper insights into narrative evolution. Although Roberta delivers high accuracy, it is computationally intensive due to its large number of parameters and deeper architecture.

3.2 Proposed algorithm

The following proposed method is shown step-by-step below for the identification and categorization of fake news.

Proposed Algorithm: for Detection and Classification of Fake News in Social Media

Step 1: Data Collection

- Acquire the PolitiFact and Gossip Cop dataset from GitHub repository, which includes tweet IDs, news URL, title, title id and label.

Step 2: Data Preprocessing

- Data Labeling, Handle Missing Values, Convert Text to Lowercase, Remove URLs, HTML Tags, Punctuation, Extra Spaces, Stop Words, Normalization etc.

Step 3: Perform exploratory data analysis (EDA) on cleaned dataset.

Step 4: Train-Test Split

- Training (80%)

- Testing (20%)

Step 5: Implement classification models, such as Roberta, on the training set.

Step 6: Train the model with hyperparameters.

Step 7: Analyze the trained models' performance using measures such as accuracy, F-score, precision, and recall.

Step 8: Final Outcome

- The best model should be deployed and tested on new, unseen data to ensure its effectiveness in real-world circumstances.

Finish!!!

CHAPTER IV:

RESULTS ANALYSIS

4.1 Experimental Configuration

A 64-bit version of Windows 10, 16 GB of RAM, 500 GB of solid-state drive (SSD), and an Intel Core i5 processor are all components of the experimental configuration used in this study. Python and its popular libraries, including Pandas, NumPy, plotly, matplotlib, seaborn, and NLTK, as well as Jupyter Notebook, are utilized in the research.

Programming Language Python

Programming with Python allows for high-level, general-purpose interpreting. Python's design philosophy prioritizes code readability through its prominent use of indentation. Programmers may use its object-oriented approach and language features to create logical code for both small and big projects (Patkar *et al.*, 2022). Python sorts its variables dynamically and collects their rubbish. Structured (especially procedural), object-oriented, and functional programming paradigms are all supported. The extensive standard library that comes with Python makes it a "batteries included" language.

There are now several implementations of Python available. One of them is Jython, which is programmed in Java for the Java Virtual Machine. Another is Iron Python, which is written in C# for the Common Language Infrastructure. Finally, there is a PyPy version that is written in R Python and translated into C. The most common and default Python implementation is C Python, which is created by the Python Software Foundation and implemented in C (Lakshmi, 2018). Though developed in their native tongue, some implementations may interact with various languages using modules. The community development model powers most of these modules, which are free and open-source. The combination of characteristics that make Python superior to other languages results in its wide range of applications.

Python is faster and more productive than ever before because to its clean object-oriented architecture, improved process management, robust integration and text processing capabilities, and built-in unit testing framework. When developing sophisticated applications for multi-protocol networks, it is seen as a practical choice(Chandel *et al.*, 2022). Since Guido Van Rossum created Python in 1991, it has undergone tremendous change. To put it briefly, it is a high-level, dynamic, interpreted programming language that makes creating a wide range of applications easier. Because of its simpler syntax and reduced learning curve, it's also simple to get started. Additionally, it is a programming language that can be used for a wide range of tasks, including embedded applications, data science, machine learning, web applications, video games, and much more.

Python Libraries

This section presents an overview of the proposed python library that has been utilized in the development of the fake news detection system.

1) Pandas

Pandas is a widely used package for working with and analyzing data. The platform provides Data Frames together with other data structures which simplify the handling of structured datasets (Rajathi M, 2021). Data scientists use Pandas due to its complete array of capabilities which support data cleaning and transformation along with investigative tasks. It also includes the necessary data structure and methods for working with numerical tables and time series.

2) NumPy

Python offers the NumPy library as its core language component for scientific computing needs. As a result, linear algebra, random number processing, and Fourier transformations are all made possible. The programming language keeps matrices and multi-dimensional arrays accessible through its library of detailed mathematical functions (Bhat, 2023).

3) Scikit-Learn

Scikit-Learn functions as an open-source Python tool that serves multiple applications including data science, data mining, analysis and ML (Khadka, 2019). The sklearn package provides numerous ML algorithms and tools which support functions such as estimation and pre-processing and splitting.

4) Matplotlib

The Python package Matplotlib enables users to generate static and interactive two-dimensional charts which work across multiple system platforms. Matplotlib provides tools for generating various plots including bar charts, histograms, error charts, power spectra as well as scatter plots and several other types. There are a number of methods to personalize the plots, such as adding animations, subplots, and 3D views, and even making them interactive (Pérez-Rosas *et al.*, 2018).

5) Seaborn

Users can generate complex data visualizations with Seaborn because this data visualization package extends Matplotlib to create a functional data visualization interface. It may be utilized for exploratory data analysis in a number of ways, such as making heatmaps, violin plots, data regression plots, and other similar approaches (Sundaram *et al.*, 2023).

6) Plotly

Plotly serves as an open-source Python module that helps users create various graphs among line charts as well as scatter plots and bar charts and histograms and area charts. Being more interactive is a benefit that it provides us. Other scripting languages can access the graphs saved in JSON data format. You may use Plotly to make charts both online and offline (Belorkar *et al.*, 2020).

7) NLTK

Python users have access to an NLTK package that provides multiple methods to process natural language. This software is user-friendly and available as open source. The platform features multiple pre-processing tools among which users can find word count functionality with tokenization features and punctuation normalization along with stop word filtering capabilities. To better comprehend, evaluate, and pre-process the text samples, NLTK is a great assistance to the computer (Navlani, 2019).

Jupyter Notebook

Jupyter Notebook is a web-based program that facilitates code creation, editing, and execution. The program may run locally, even without an internet connection, or it can be uploaded to a distant server and run there (Fu and Jiang, 2019). Jupyter Notebook may be used with Anaconda by opening the application on the project home page. This will create a new browser window with a control panel listing Jupyter Notebook files. The kernel in the panel has the ability to open or close all local files(Prathanrat and Polprasert, 2018). Jupyter Notebook's kernel is analogous to an automobile's engine. The kernel runs the programs listed on the control panel. Launching the corresponding kernel is an automated process that occurs when opening a code file. Finally, the Terminal, Workbench, and Viewer programs all include file managers that can be used to manage all the relevant files (also known as "notebooks"). These files contain code and a lot of text components that are created by Jupyter Notebook(Smajić, Grandits and Ecker, 2022).

4.2 Performance Measure

An essential component of developing a successful machine learning model is doing model evaluations. This section presents the various performance indicators that have been employed to verify the effectiveness of research work(Shoemaker, 2019). Unquestionably, quantitative metrics are necessary to assess the applicability and accuracy of model while justifying the efficacy of a certain computational technique in solving a given research issue. The following are the validation measures employed in the research work:

Confusion Matrix

An ML-based classification issue with several potential output classes can be measured via a confusion matrix. This method provides a comprehensive evaluation of a model's performance by comparing its predictions to the ground truth labels in a given dataset (Alarfaj and Khan, 2023). A square table is the standard format for the matrix, with rows representing the actual classes and columns representing the expected classes. The following figure 4.1 presents the visual representation of confusion matrix.

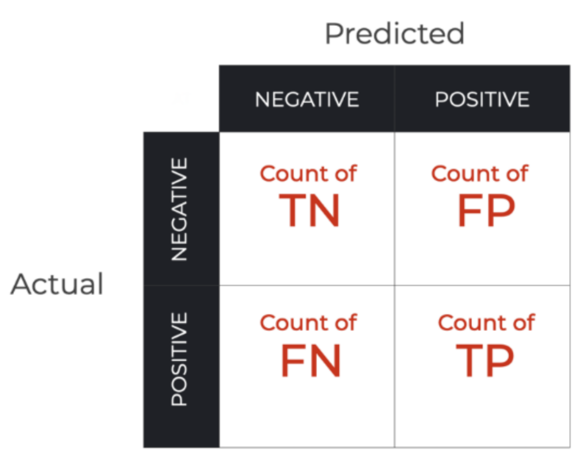

Figure 4.1: Confusion Matrix

- **Accuracy:** The entire evolution of the model may be characterized by its accuracy. The accuracy rate is the percentage of times a data point is correctly classified by the algorithm. The basic formula is provided by Equation (4.1), which entails dividing the total number of identified data points by the fraction of properly categorized data points.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \dots\dots\dots\dots\dots (4.1)$$

- **Precision:** The accuracy with which a prediction is made with respect to a set of positive data is called precision (Mishra and Sadia, 2023), which is shown in Equation (4.2).

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots\dots\dots\dots\dots\dots... (4.2)$$

- **Recall:** The capacity of the algorithm to identify all instances of fake news is measured by recall. It determines the percentage of real fake news items out of all the fake news pieces, including the ones that were overlooked (false negatives, FNs). The bulk of fake news stories will be recognized due to high recall, which is essential for reducing the possibility of missed incorrect information (Al-tarawneh *et al.*, 2024). It can be evaluated using Equation (4.3).

$$Recall = \frac{TP}{TP + FN} \dots\dots\dots\dots\dots\dots\dots\dots\dots.. (4.3)$$

- **F1-Score:** A balanced metric that is particularly helpful in cases of class imbalance is the F1-score, which is the harmonic mean of precision and recall (Moisi *et al.*, 2024). It can be evaluated using Equation (4.4).

$$F1 - Score = \frac{2 * (Precision * Recall)}{Precision + Recall} \dots\dots\dots\dots.. (4.4)$$

- **ROC Curve:** AUC-ROC is a binary classification assessment approach that assesses a classifier's ability to discriminate between classes at different threshold values(Mohsen *et al.*, 2024). As the area under the curve (AUC) grows, the model's class predictions become more accurate. One way to generate a ROC curve is by calculating the True Positive Rate (TPR) and the False Positive Rate (FPR).

4.3 Error Analysis

Error analysis uncovered several types of challenging cases where the model struggled. These included:

- **Sarcasm and Satire:** Articles using sarcasm or humor were often misclassified due to their subtle cues and figurative language. Roberta, lacking real-world knowledge and tone comprehension, frequently interpreted sarcastic content as literal.
- **Ambiguous Phrasing:** Headlines and articles with vague language or misleading sentence structures led to confusion in classification. Phrases without clear context or those designed to provoke curiosity often misled the model.
- **Regional Dialects and Slang:** Informal expressions, idioms, or non-standard grammar commonly used in specific regions reduced the model's accuracy. These linguistic variations deviate from the training distribution and can obscure the true intent.

For example, a satirical article titled “Aliens Endorse Presidential Candidate” was incorrectly flagged as real due to the serious tone and structured formatting mimicking legitimate news. Similarly, a colloquial headline like “He got the boot!” was misclassified as fake news because the slang term ‘got the boot’ was interpreted literally. These examples highlight the limitations of the model’s semantic and cultural understanding, suggesting the need for integrating context-aware modules or external knowledge bases to enhance comprehension of nuanced language.

4.4 Dataset Description

In spite of the abundance of publicly available datasets pertaining to the identification of false news across several domains, the Fake News Net dataset obtained from a GitHub repository was used in this investigation. Labelled news articles from two websites—politifact.com9 and gossipcop.com—are included in the collection. From now on, we will refer to them as PolitiFact and Gossip Cop. The collection includes graphical and textual information, all tweets and retweets for each news item, and user details for the relevant Twitter accounts for each item. The separate description of the PolitiFact and Gossip Cop datasets are given below to properly understand them.

Description of Proposed PolitiFact Dataset

The PolitiFact dataset within Fake NewsNet presents 762 records organized across 5 fields which include id, news_url, title, tweet_ids along with label. News trustworthiness assessments in this dataset incorporate both news content elements like title and URL and social media information conveyed through tweet IDs as part of the analysis. The label column functions as the target variable, facilitating classification or analytical operations to detect bogus news. The dataset enables false news identification applications and analyses news distribution dynamics on Twitter while studying public news interactions through integrated article information and social media activities.

Visualization Insights for the PolitiFact Dataset

This section presents the visualization results for the PolitiFact dataset to completely understand the dataset and its hidden pattern for building an efficient fake news detection system.

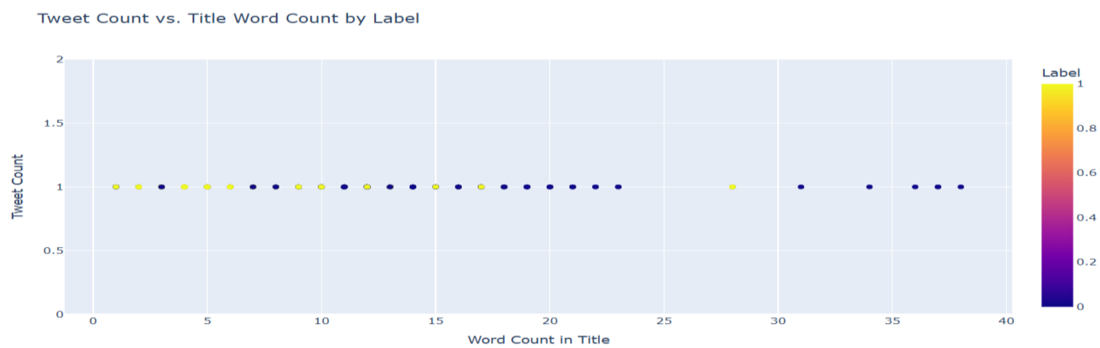

Figure 4.2: Scatter Plot for Distribution of Tweet Count vs. Title Word Count by Label on PolitiFact Dataset

The relationship between word counts in article titles and tweet counts from PolitiFact appears in Figure 4.2 as a scatter plot. The graph's x-axis shows the word count in the title ranging between 0 and 40, while a y-axis indicates a tweet count ranging between 0 and 2. The colour gradient, from purple (label 0) to yellow (label 1), represents the classification of news articles. The data points mostly reveal a tweet count of 1 across diverse title word counts, indicating no fluctuation in tweet activity. The visual suggests that title length does not significantly influence tweet counts for the dataset, as most articles cluster at a single tweet irrespective of their label. (Word count: 100).

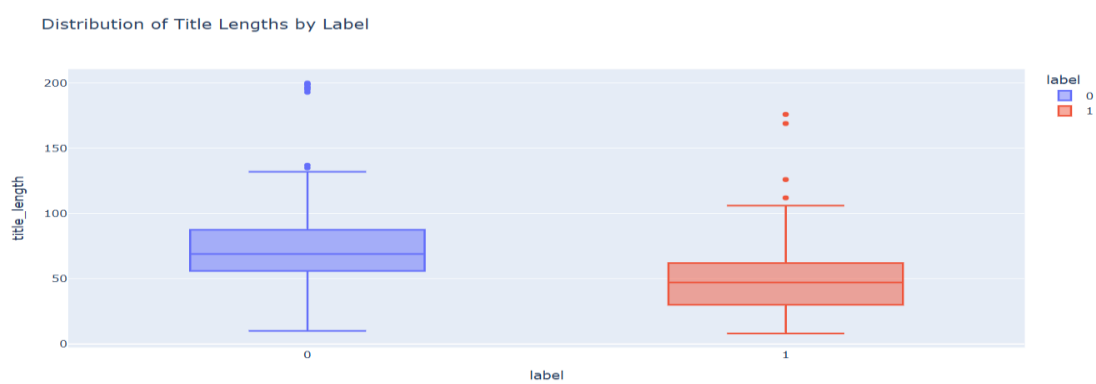

Figure 4.3: Box Plot for Distribution of Title Lengths by Label on PolitiFact Dataset

Figure 4.3 displays a box plot that depicts the distribution of title lengths categorized by label within the PolitiFact dataset. The title lengths for label 0 (blue) show a median of about 50, accompanied by a broader distribution and many outliers beyond

150 characters. Conversely, the title lengths for label 1 (red) have a lower median, about 60, characterized by a more concentrated distribution and fewer outliers. This visualization illustrates the disparity in title length distributions between the two labels, indicating that titles labelled as 1 are often shorter and more uniform than those labelled as 0.

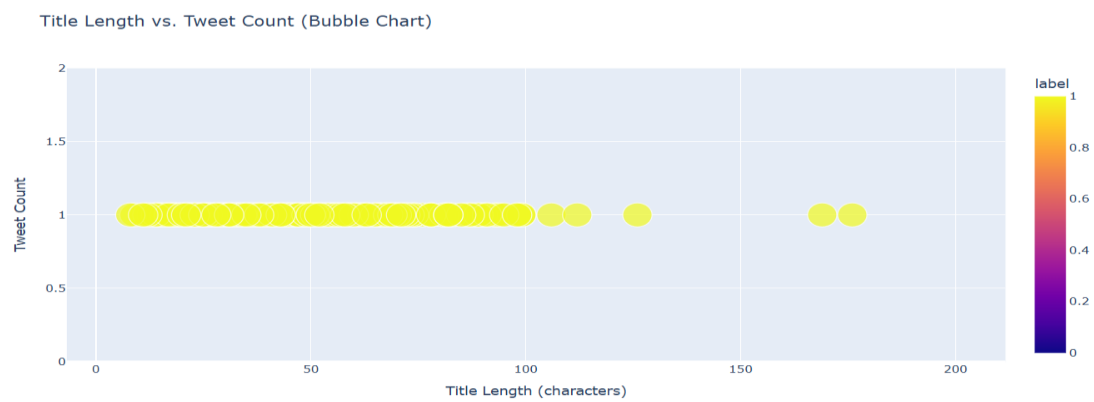

Figure 4.4: Scatter Plot for Relationship Between Title Length and Tweet Count on PolitiFact Dataset

Figure 4.4 illustrates a bubble chart depicting the relationship between title length (in characters) on the PolitiFact dataset. The graph's x-axis displays a title length ranging between 0 and 200, while the y-axis indicates a tweet count ranging between 0 and 2. Each data point represents an article, with its size proportional to a specific parameter and colour indicating the label (ranging from 0 to 1). The title lengths span from 0 to over 200 characters, while tweet counts are predominantly clustered around 1. The colour gradient (purple to yellow) represents the label classification, emphasizing limited variation in tweet count regardless of title length. This indicates the small effect of title length on tweet engagement metrics.

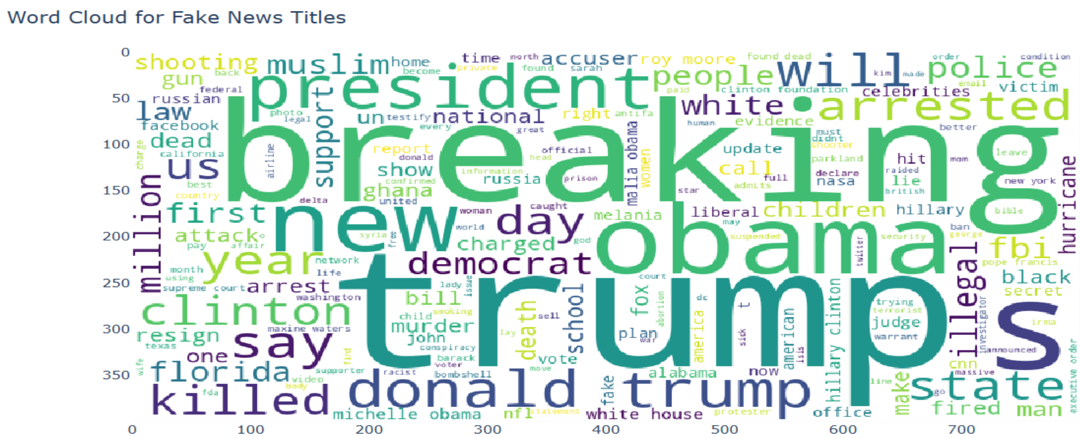

Figure 4.5: Word Cloud of Frequent Terms in Fake News Titles on PolitiFact Dataset

Word Cloud in Figure 4.5 displays the most overrepresented terms in the fake news titles of the PolitiFact dataset. A size of every word shows the frequency of the words, where big letters points to the most used words in the title. The word cloud also involves many politically sensitive words and phrases such as “president”, “trump”, “Clinton”, “Obama”, “arrested” “killed” showing that the trend of fake news titles are related political figures and cataclysm events. Such a type of visualization can illustrate useful information about the language and topics used in the dataset that is related to false information or fabricated news.

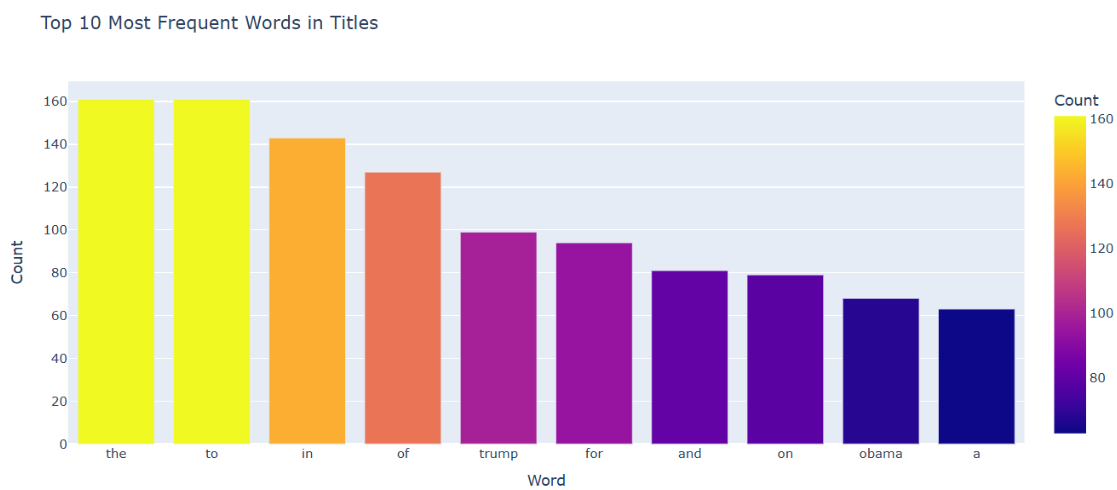

Figure 4.6: Count Plot of Top 10 Most Frequent Words in Titles on PolitiFact Dataset

Figure 4.6 illustrates the count plot of the top 10 most frequently occurring words in the titles of PolitiFact datasets. An x-axis of a graph displays the words: 'the', 'to', 'in', 'of', 'trump', 'for', 'and', 'on', 'Obama', and 'a', while, a y-axis shows a frequency of every term inside the dataset. The bar graph clearly illustrates that the word 'the' has the greatest frequency, while the term 'a' has the lowest frequency.

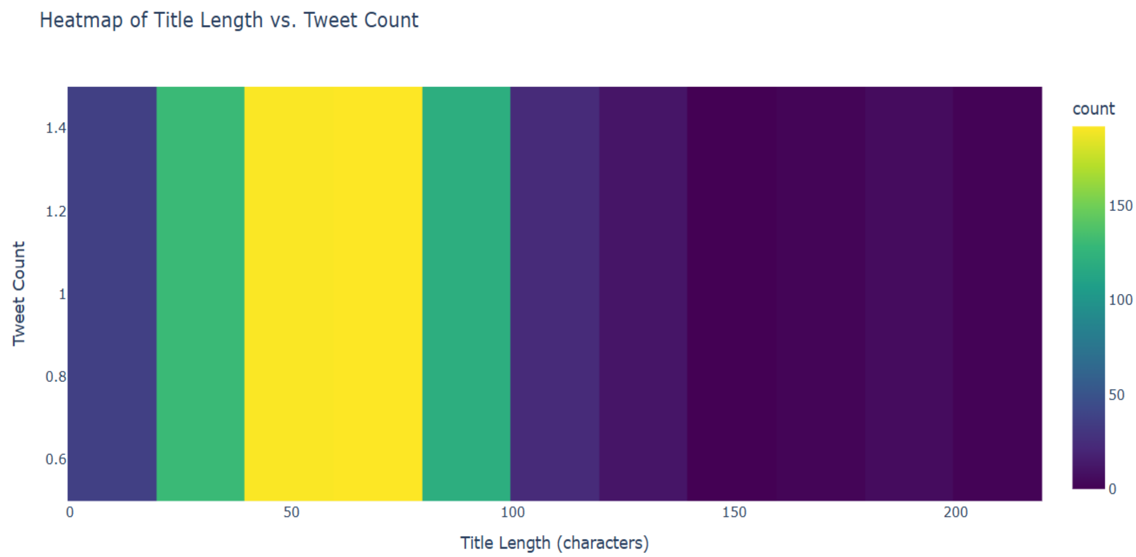

Figure 4.7: Heatmap of Title Length vs. Tweet Count on PolitiFact Dataset

Figure 4.7 is a heatmap of the PolitiFact dataset showing the correlation between title length (in characters) and tweet count. The horizontal axis measures the title length from 0-200 characters while the vertical axis measures the tweet count. The heatmap uses a colour scale to represent the density or frequency of the data points with high density having darker colours while points of lower density have lighter colours. A trend that can clearly be identified from the heatmap is that titles that are between 50 and 100 characters long are likely to have greater tweet counts than titles that are less than 50 characters or more than 150 characters.

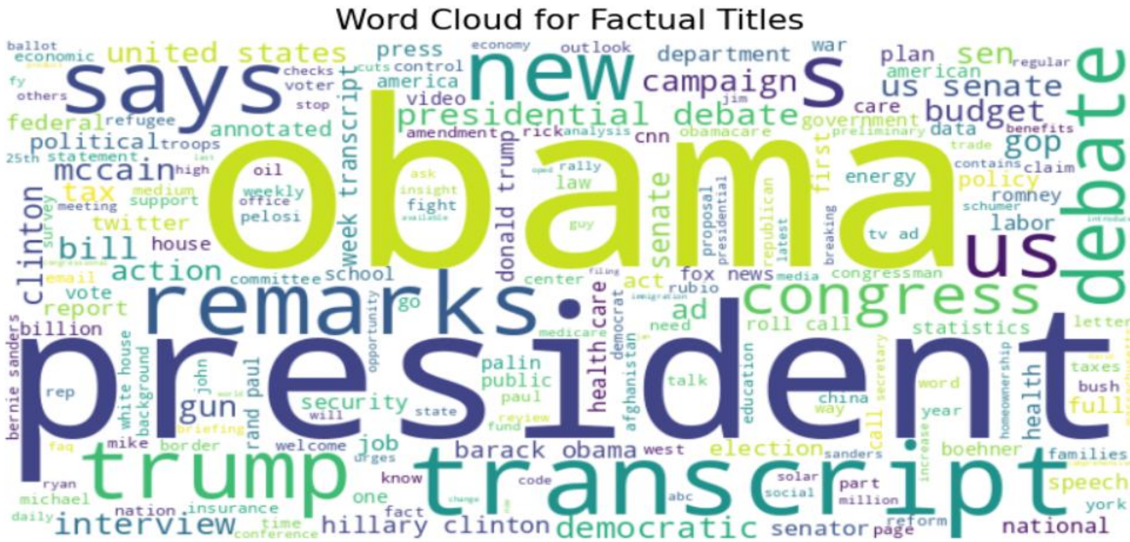

Figure 4.8: Word Cloud for Factual Titles on PolitiFact Dataset

This visualization of factual titles from PolitiFact shows word frequency using a word cloud format as presented in Figure 4.8. In the figure, a size of every word indicates its frequency, with more often occurring words shown in bigger font inside the cloud. The compact visual portrayal through word cloud summarizes main topics from factual titles that include "president," "Congress," "states," "campaign" and "election." This word cloud utilizes graphical techniques to display the key concepts found in the factual titles within the PolitiFact data collection.

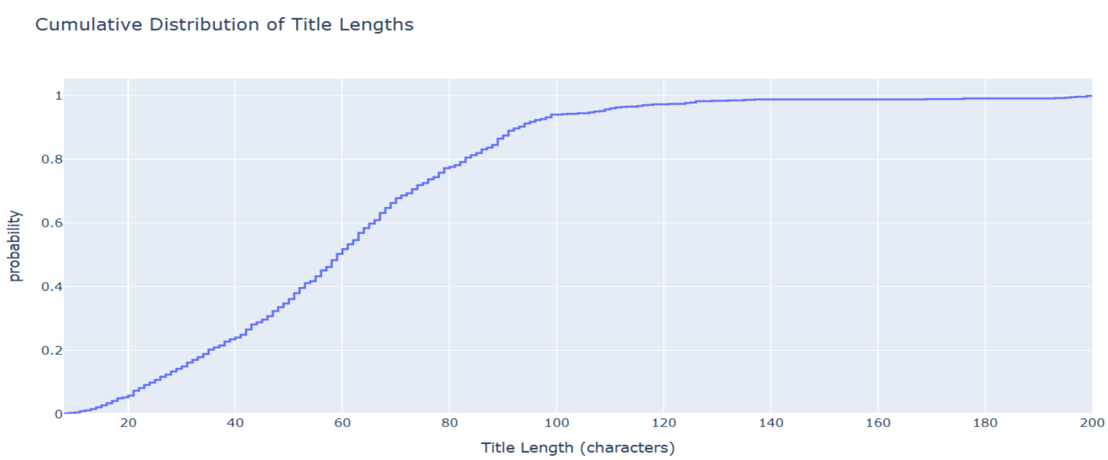

Figure 4.9: Line Graph for Cumulative Distribution of Title Lengths on PolitiFact Dataset

The distribution chart for title lengths across the PolitiFact database appears in Figure 4.9. In terms of cumulative probability, the graph shows the following values: 20-character titles = 0.1, 40-character titles = 0.3, 60-character titles = 0.5, 80-character titles = 0.65, 100-character titles = 0.8, 120-character titles = near 0.9, 140-character titles = at 0.95, 160-character titles = around 0.98, 180-character titles = about 0.99-, and 200-character titles = close to 1.0. The distribution of title lengths within this dataset becomes easier to grasp through this visualization because it supports modeling approaches while showing the typical features of news articles.

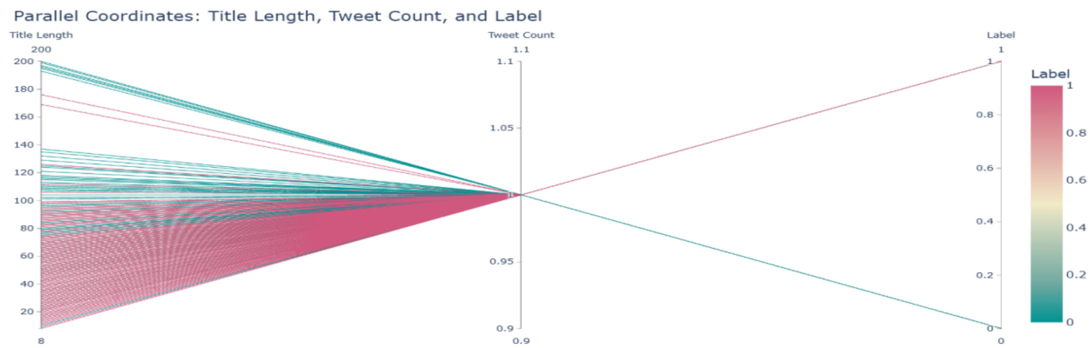

Figure 4.10: Parallel Coordinates: Title Length, Tweet Count, and Label on PolitiFact Dataset

Figure 4.10 presents a parallel coordinates plot showing the relationship amongst three variables from the PolitiFact dataset which comprise title length and tweet count and label. The title length is shown on the left axis, ranging from 8 to 200 characters. The tweet count is displayed in the middle axis, ranging from 0.9 to 1.1. The label is represented on the right axis, with values from 0 to 1. Every data point on the graph uses lines between axes to present values which help viewers recognize patterns that exist between the three variables.

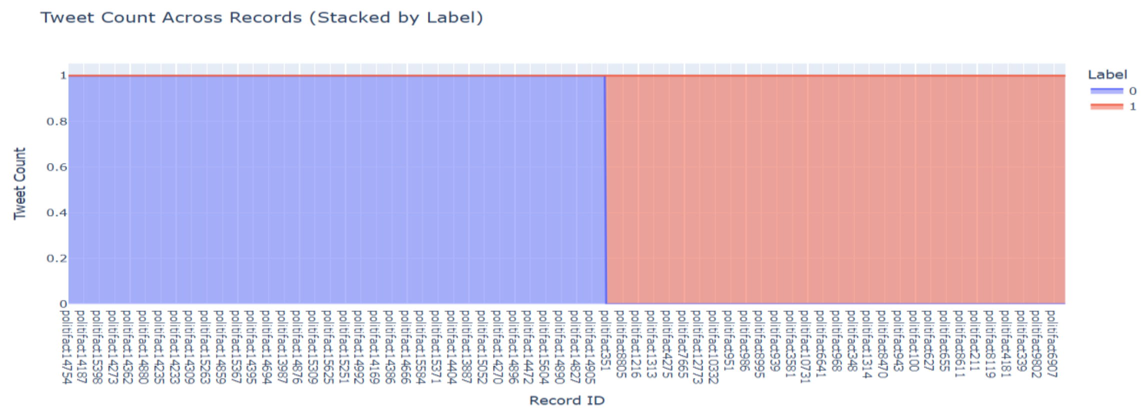

Figure 4.11: Area Chart of Tweet Count Across Records (Stacked by Label) on PolitiFact Dataset

The graphic representation of tweet frequency in the PolitiFact records appears as a stacked area chart in Figure 4.11. The records are arranged according to their label (0 or 1). The tweet count is depicted on the y-axis, while the record ID is represented on the x-axis. The records with label 0 are represented by the blue area, while the records with label 1 are represented by the orange area. The distribution of tweet counts across the various records in the dataset and the variation in tweet counts between the labelled categories can be analyzed using this visualization. This information is beneficial for comprehending the social media engagement patterns associated with the news articles in the PolitiFact dataset.

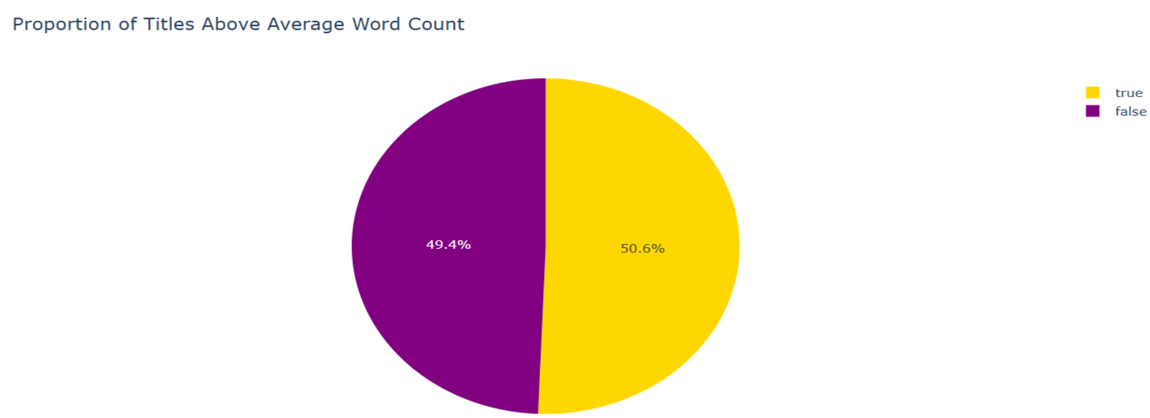

Figure 4.12: Pie Chart for Proportion of Titles Above Average Word Count on PolitiFact Dataset

Figure 4.12 shows a break-up of news article titles through the pie chart to demonstrate word count proportions between true and false content. This chart identifies 50.6% of articles fall within the true category while false content makes up 49.4%. This almost equal distribution indicates that the categorization of news stories as true or false is not much affected by the presence of above-average word counts in their titles.

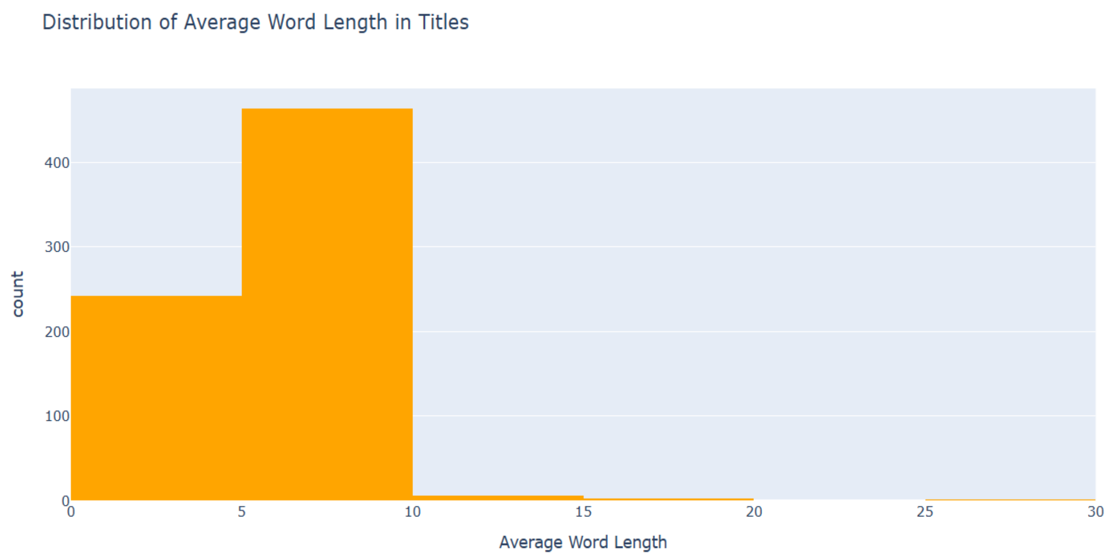

Figure 4.13: Histogram for Distribution of Average Word Length in Titles on PolitiFact Dataset

Figure 4.13 illustrates the histogram for the distribution of the average word length in titles on the PolitiFact dataset. The y-axis displays the total number of titles falling into each word length group, while the x-axis shows the average word length, which can be anywhere from zero to thirty characters. The distribution is heavily skewed, with the majority of titles having an average word length between 0 and 5 characters, as seen by the tall bar in the lower left. Fewer titles show longer average word lengths, with the count tapering off as the word length increases.

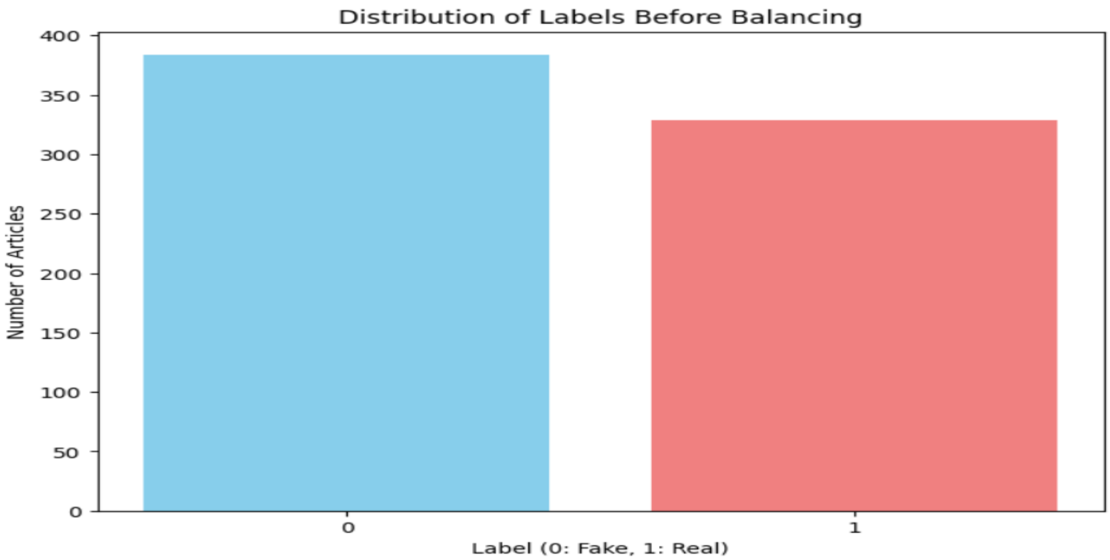

Figure 4.14: Bar Graph for Distribution of Labels Before Balancing on PolitiFact Dataset

Figure 4.14 shows a bar graph that illustrates the distribution of labels prior to balancing on the PolitiFact dataset. There are two classes on the x-axis: 0 (fake) and 1 (actual). Article counts for each category are shown on the y-axis. Clearly, the dataset is skewed towards the "real" category (383 articles) rather than the "fake" category (333 articles), as seen in the graph.

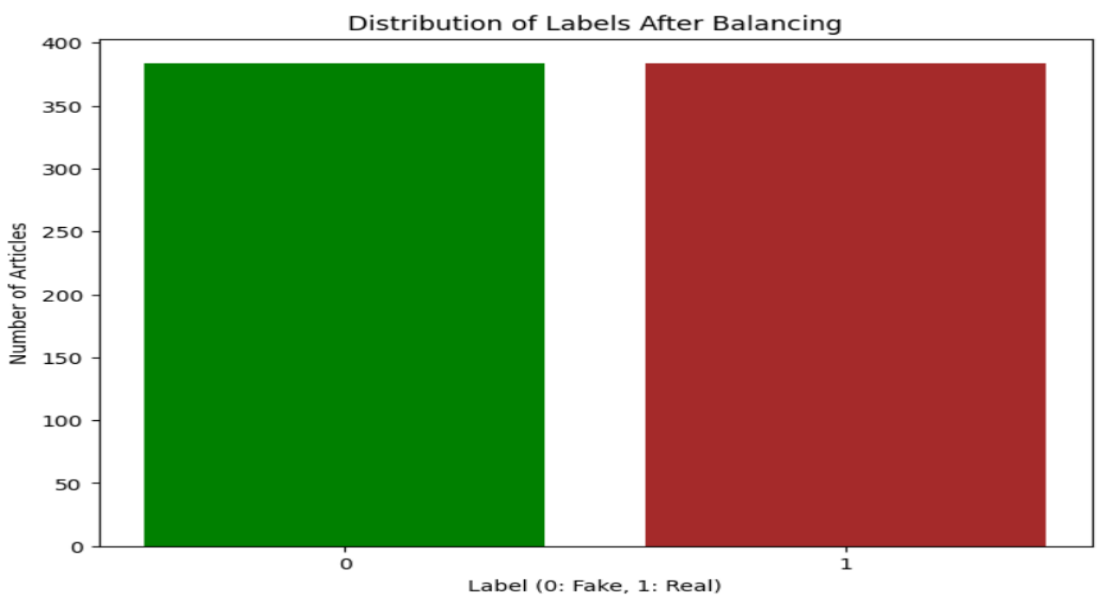

Figure 4.15: Bar Graph for Distribution of Labels After Balancing on PolitiFact Dataset

The distribution of labels in the PolitiFact dataset following data balancing using the SMOTE is shown in Figure 4.15. The bar chart shows the number of articles categorized as "Fake" (label 0) and "Real" (label 1). Before SMOTE, there was an imbalance with significantly fewer "Fake" articles. After SMOTE, the distribution is more balanced, with approximately equal numbers of "Fake" and "Real" articles, indicating that SMOTE has successfully addressed the class imbalance issue.

Description of Proposed Gossip Cop Dataset

The Gossip Cop dataset, also a component of the Fake NewsNet collection, has 20,465 entries and 5 attributes: id, news_url, title, tweet_ids, and label, aimed at assessing news reliability. Each entry denotes a news article with distinct identifiers (id) such as "gossipcop-1697863049," metadata including the article's URL (news_url) and headline (title), social media context through tab-separated tweet IDs (tweet_ids), and a binary classification label (label) indicating the authenticity of the article, with 0 representing fake and 1 representing real. The dataset combines article-level information with social media interactions, making it suitable for applications such as false news identification, study of news dissemination on social media, and assessment of public involvement with news stories.

Visualization Insights for the Gossip Cop Dataset

This section presents the visualization results for the Gossip Cop dataset to completely understand the dataset and its hidden pattern for building an efficient fake news detection system.

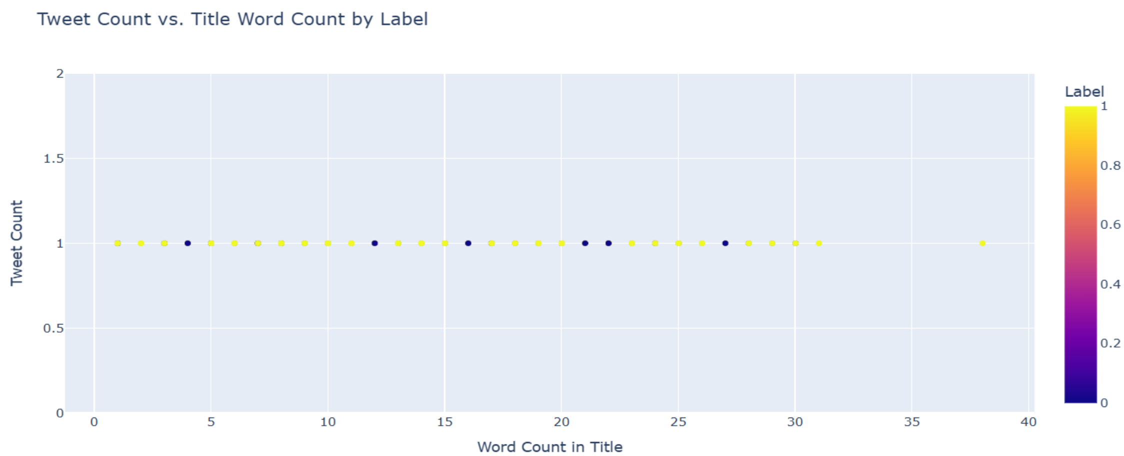

Figure 4.16: Scatter Plot for Distribution of Tweet Count vs. Title Word Count by Label on GossipCop Dataset

Figure 4.16 shows a scatter plot illustrating the distribution of tweet counts based on the number of words in their titles for the GossipCop dataset. A single point on the visualization represents a tweet while the color spectrum identifying each point indicates whether the news is considered fake (blue) or real (yellow). The x-axis ranges from 0 to 40, representing the word count in the title, while the y-axis ranges from 0 to 2, representing the tweet count. The plot reveals that the majority of tweets, regardless of their label, have a title word count between 0 and 10, with a corresponding tweet count of 1.

Figure 4.17: Box Plot for Distribution of Title Lengths by Label on GossipCop Dataset

Figure 4.17 is a box plot displaying the typical length of titles for "Fake" and "Real" news items from the Gossip Cop dataset. The x-axis uses label, 0 as Fake and 1 as Real to classify the data in the given graph. On the y-axis, one gets the title length. A comparison of the box plots shows that the median length of titles of "Fake" news is about 60 words, while that of "Real" news is about 80 words. The length of the box which represents the interquartile range (IQR), is also less for "Fake" news when compared to the "Real" news. Further, the whiskers also show the corresponding range of length of titles, where it is evident that titles of "Fake" news have higher variability compared to "Real" news.

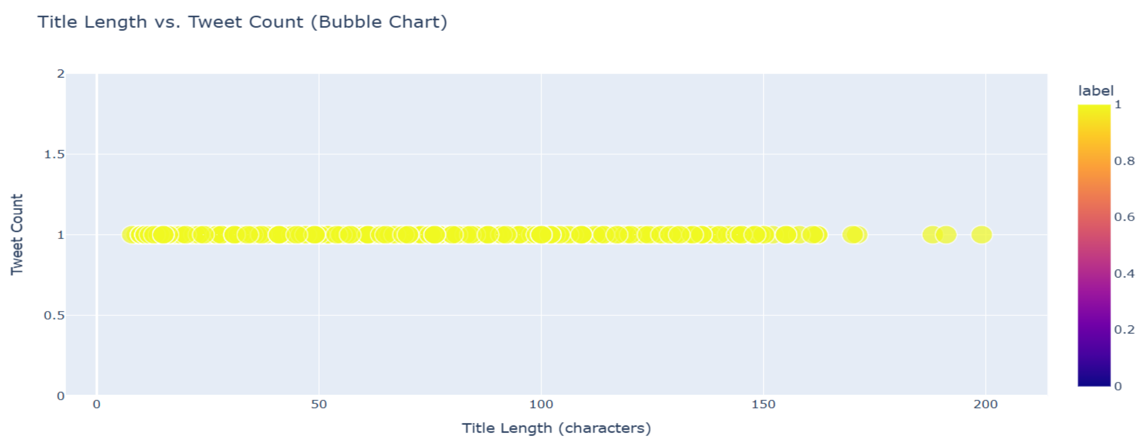

Figure 4.18: Scatter Plot for Relationship Between Title Length and Tweet Count on GossipCop Dataset

Figure 4.18 presents a bubble chart illustrating the relationship between title length and tweet count for the Gossip Cop dataset. The title length in characters, which ranges from 0 to 200, is shown by the x-axis. The tweet count, which ranges from 0 to 2, is displayed by the y-axis. Each bubble represents a cluster of tweets with similar title lengths and tweet counts. The color gradient signifies the label, with blue indicating "Fake" news and yellow indicating "Real" news. The plot reveals that the majority of tweets, regardless of their label, have a title length between 0 and 50 characters and a tweet count of 1. Additionally, there is a cluster of tweets with a title length between 150 and 200 characters and a tweet count of 1.

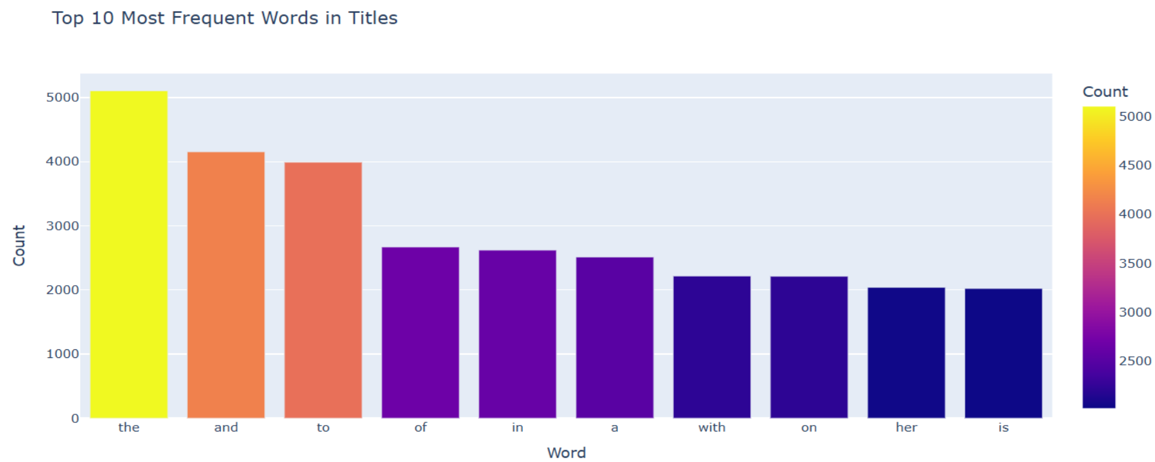

Figure 4.19: Count Plot of Top 10 Most Frequent Words in Titles on GossipCop Dataset

Figure 4.19 shows a bar graph illustrating the ten most prevalent terms in titles from the Gossip Cop dataset. An x-axis presents these terms, while a y-axis shows their corresponding frequencies. With almost 5,000 occurrences, the plot indicates that "the" is the most common word. The words "and" and "to," which appear 4000 times each, follow shortly behind. The frequency with which the remaining words, including "of," "in," "a," "with," "on," "her," and "is," occur decreases.

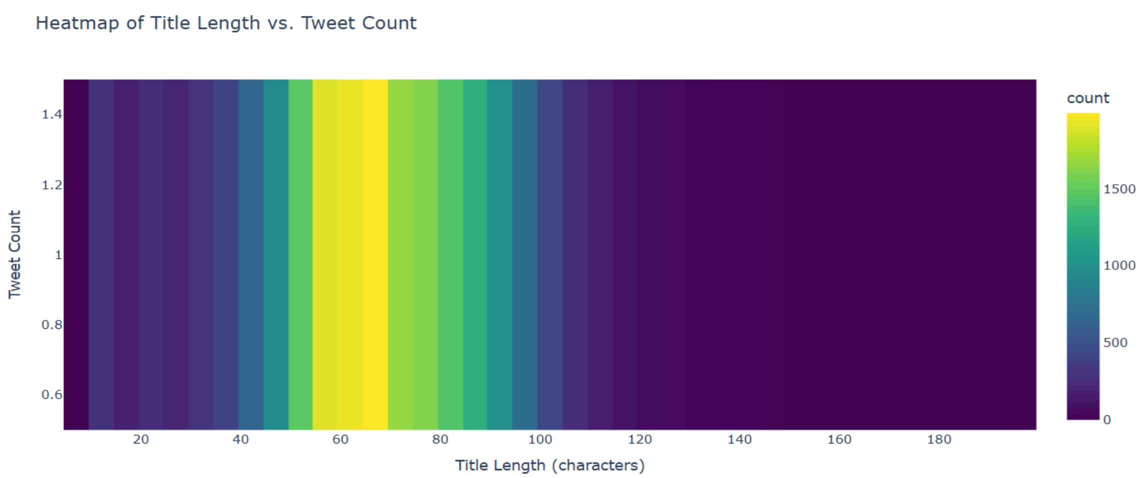

Figure 4.20: Heatmap of Title Length vs. Tweet Count on GossipCop Dataset

Figure 4.20 presents a heatmap illustrating the relationship between title length and tweet count for the Gossip Cop dataset. An x-axis displayed the title length in characters, ranging between 0 and 180. A y-axis shows the tweet count, ranging between 0.6 and 1.4. The color intensity represents the frequency of tweets with a specific combination of title length and tweet count. The heatmap reveals that the majority of tweets have a title length between 50 and 80 characters and a tweet count of 1. Additionally, there is a cluster of tweets with a title length between 10 and 20 characters and a tweet count of 1.2.

Figure 4.21: Word Cloud for Factual Titles on GossipCop Dataset

Figure 4.21 presents the word cloud of identified fact-based articles' titles in the context of the Gossip Cop dataset containing the most frequently used primary words. A size of every word indicates the frequency of the words used, the larger the word the more frequent it was used. Kim Kardashian, Taylor Swift, Meghan Markle, wedding, love, baby, family, reveal, look and season are some of the standout words in the corpus. This word cloud summarizes the main topics that factual news articles in the Gossip Cop dataset are more likely to address.

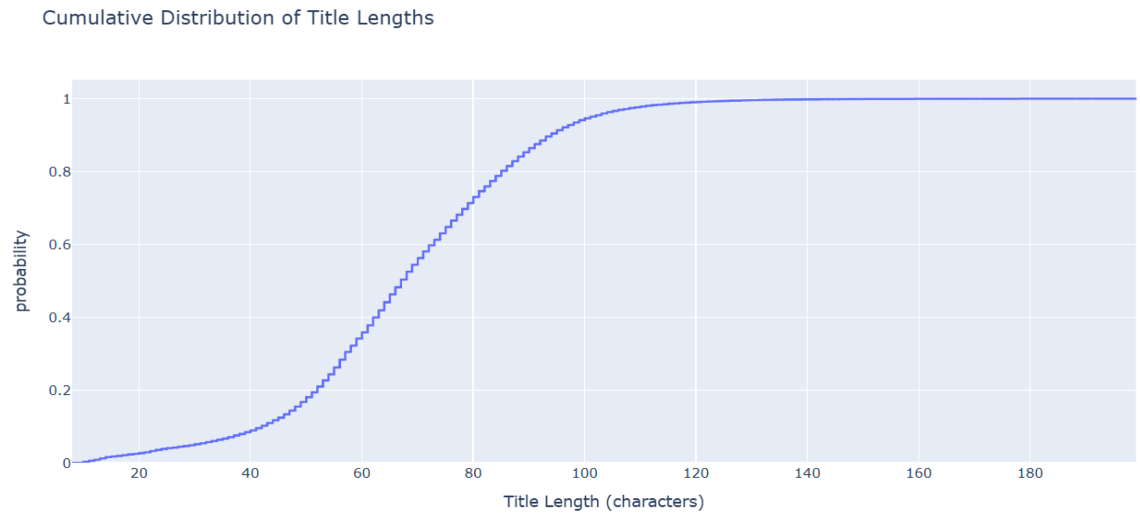

Figure 4.22: Line Graph for Cumulative Distribution of Title Lengths on GossipCop Dataset

Figure 4.22 presents a line graph illustrating the cumulative distribution of title lengths for the Gossip Cop dataset. The title length in characters, which ranges from 0 to 180, is shown by the x-axis. The cumulative probability, which runs from 0 to 1, is represented by the y-axis. The graph shows that the majority of titles have a length between 50 and 100 characters. The cumulative probability reaches 1 around a title length of 180 characters, indicating that almost all titles are shorter than 180 characters.

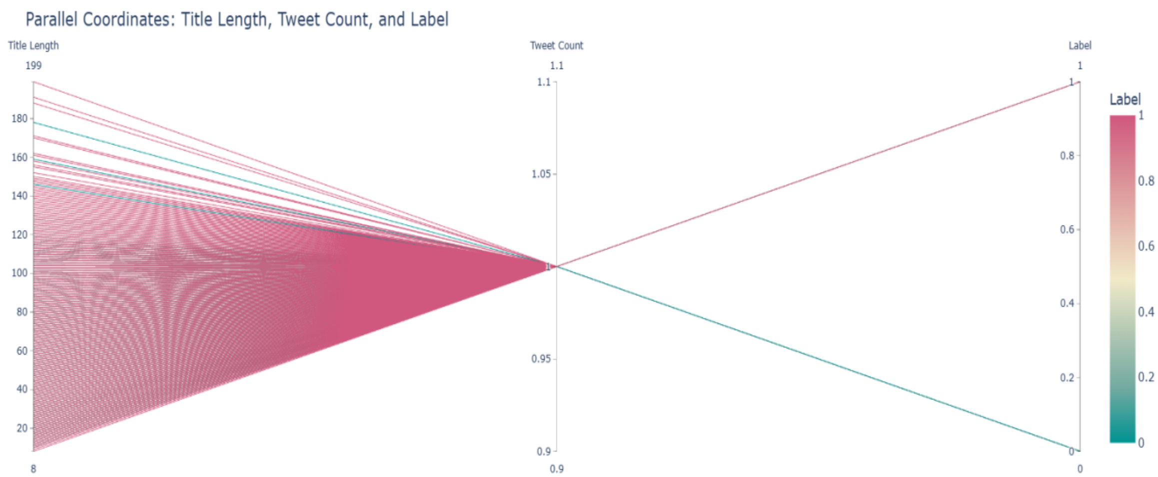

Figure 4.23: Parallel Coordinates: Title Length, Tweet Count, and Label on GossipCop Dataset

Figure 4.23 presents a parallel coordinate plot illustrating the relationship between title length, tweet count, and label in the Gossip Cop dataset. Each line represents a tweet, and the colour gradient signifies the label, with blue indicating "fake" news and yellow indicating "real" news. The x-axis represents the three variables: title length (ranging from 0 to 199 characters), tweet count (ranging from 0.9 to 1.1), and label (0 for "fake" and 1 for "real"). The plot reveals that the majority of tweets have a title length between 50 and 100 characters, a tweet count of 1, and are labeled as "real" news. Furthermore, a group of tweets characterized by titles ranging from ten to twenty characters in length and a total of twelve tweets are being referred to as "fake" news.

Figure 4.24: Area Chart of Tweet Count Across Records (Stacked by Label) on GossipCop Dataset

Figure 4.24 presents an area chart illustrating the tweet count across different records in the Gossip Cop dataset, with the data stacked by label. An x-axis presents the record ID, and a y-axis presents the tweet count. The blue area represents the tweet count for the label "real," and the red area represents the tweet count for the label "fake." The chart reveals that most records have a tweet count of 1, with the majority being labeled as "real" news. There are a few records with a tweet count of 0.

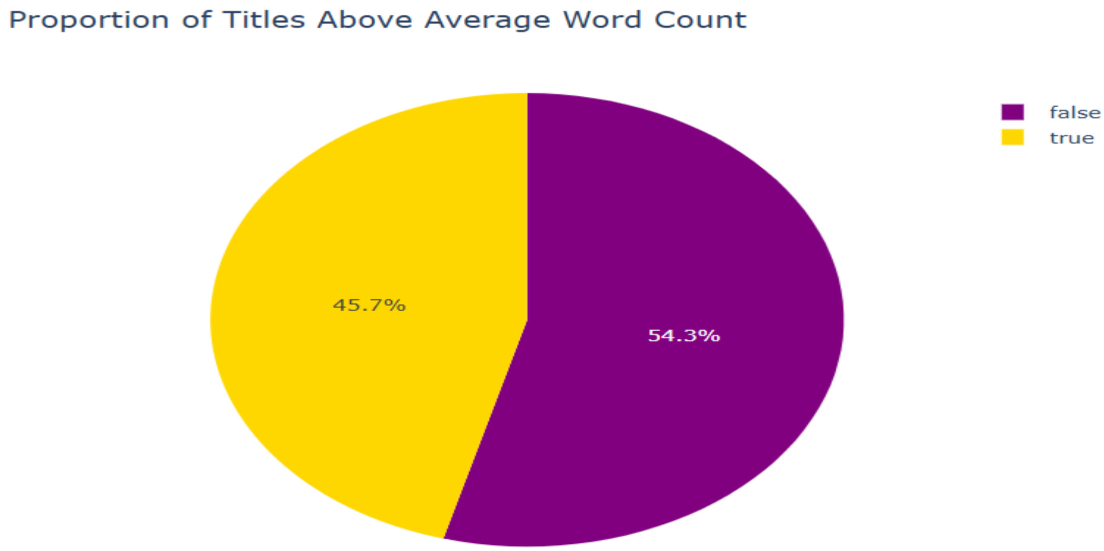

Figure 4.25: Pie Chart for Proportion of Titles Above Average Word Count in GossipCop Dataset

Figure 4.25 illustrates the percentage of titles in the Gossip Cop dataset that exceed the average word count. The chart shows that 54.3% of the titles have a word count above average, while 45.7% of the titles have a word count below average. This indicates that a slight majority of titles in the dataset have a word count higher than the average word count.

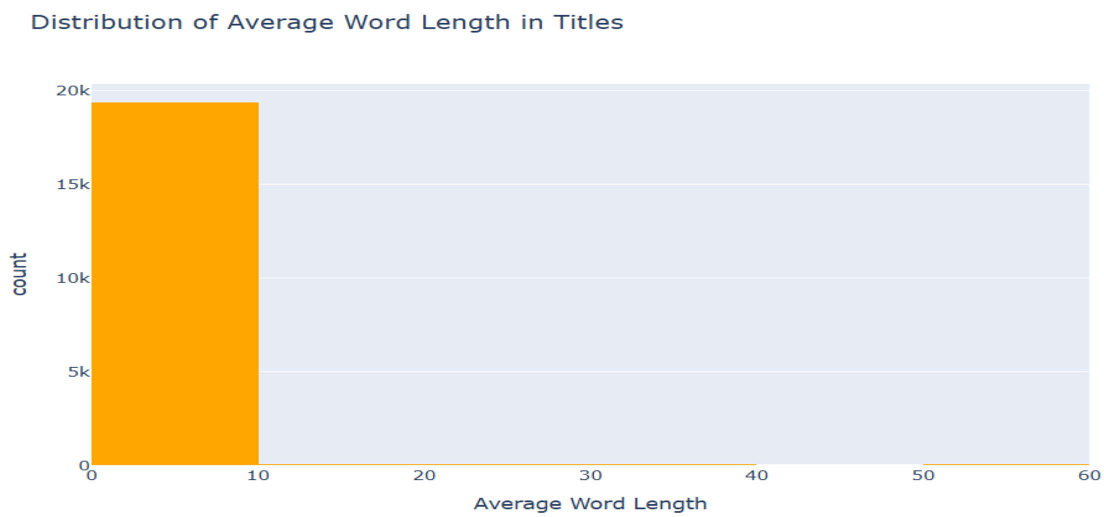

Figure 4.26: Histogram for Distribution of Average Word Length in Titles on GossipCop Dataset

Figure 4.26 illustrates the histogram for the distribution of the average word length in titles on the Gossip Cop dataset. An x-axis displays the average word length, ranging from 0 to 60. A y-axis shows the count of titles with a specific average word length. The histogram indicates that most titles possess an average word length of about 5. Moreover, there exists a reduced number of titles characterized by elevated average word lengths, indicating a right-skewed distribution.

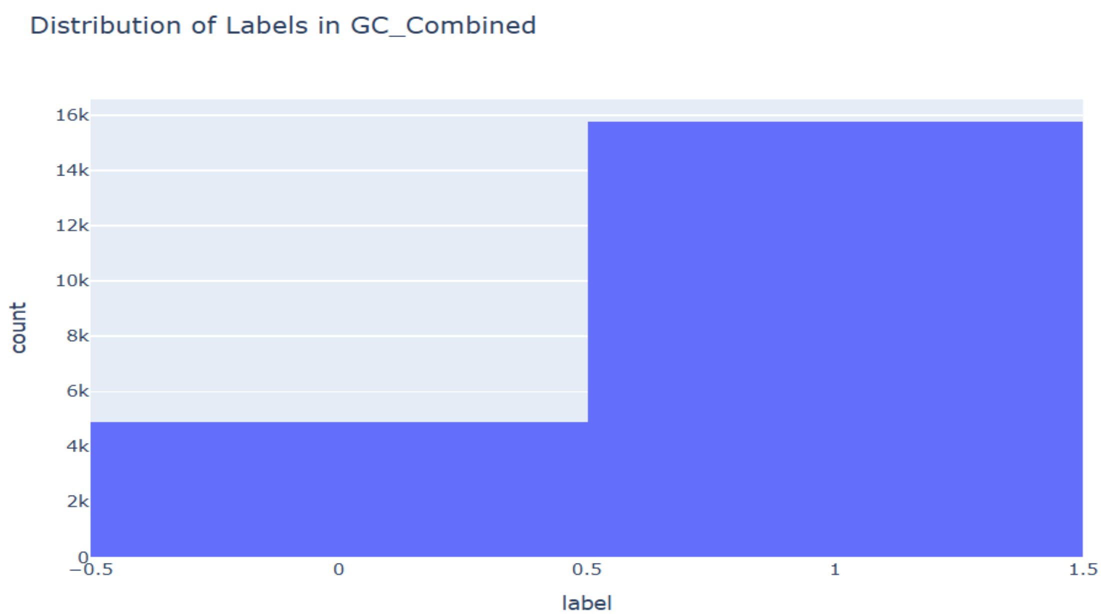

Figure 4.27: Histogram for Distribution of Average Word Length in Titles on GossipCop Dataset

Figure 4.27 illustrates the histogram for the distribution of the average word length in titles within the Gossip Cop dataset. The x-axis represents the label, with 0 indicating "fake" news and 1 indicating "real" news. The y-axis represents the count of articles with each label. The histogram shows that there are significantly more articles labeled as "real" news (approximately 15,000) compared to "fake" news (approximately 5,000), indicating a class imbalance in the dataset.

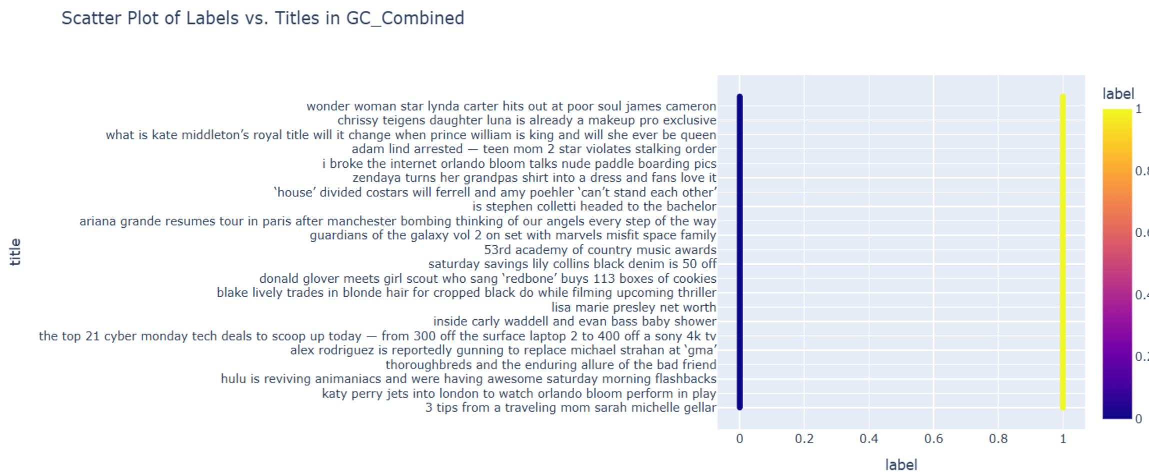

Figure 4.28: Scatter Plot of Labels vs Titles in GossipCop Dataset

Figure 4.28 presents a scatter plot illustrating the relationship between the label (indicating whether the news is "fake" or "real") and the titles of articles in the Gossip Cop dataset. The x-axis shows the label (0 for "fake" and 1 for "real") and the y-axis shows the title of the article; each point on the plot implies an article. The plot reveals that the majority of articles are labeled as "real," with a few articles labeled as "fake.". Titles range in length and subject matter, as seen by the uneven distribution of titles along the y-axis.

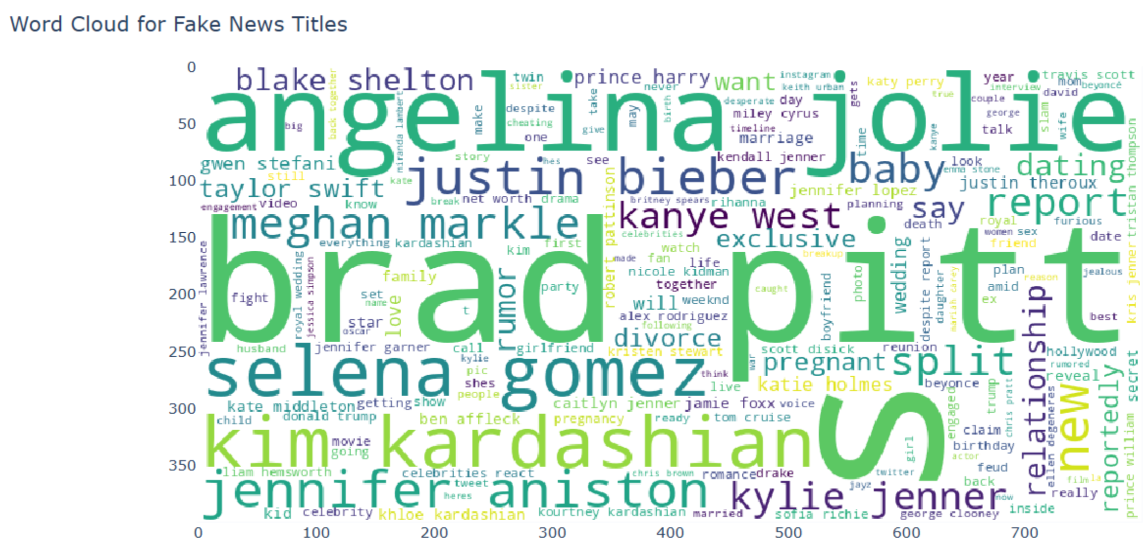

Figure 4.29: Word Cloud of Fake News Titles on GossipCop Dataset

Figure 4.29 presents a word cloud illustrating the most frequent words found in the titles of fake news articles within the Gossip Cop dataset. Each word's size is correlated with its frequency; bigger words are more likely to appear. Some of the prominent words include "Justin Bieber," "Kanye West," "Brad Pitt," "Selena Gomez," "Kim Kardashian," "Jennifer Aniston," and "Kylie Jenner." This word cloud shows the most popular subjects and themes mentioned in the Gossip Cop dataset of false news articles.

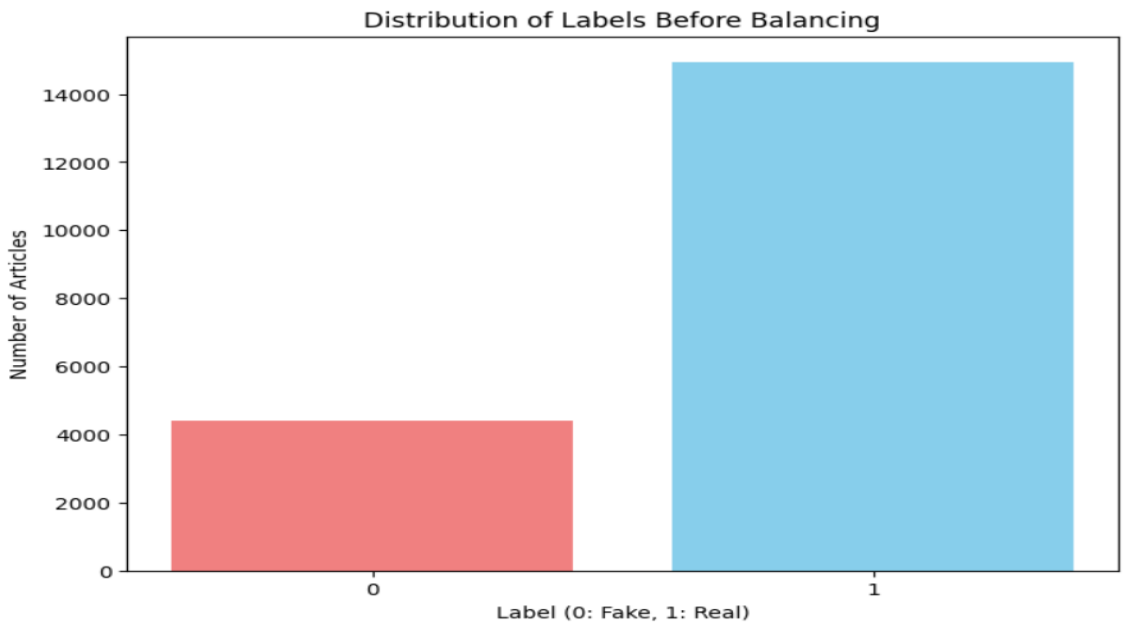

Figure 4.30: Bar Graph for Distribution of Labels Before Balancing on GossipCop Dataset

Figure 4.30 displayed a distribution of labels before balancing on the Gossip Cop dataset. The data is organized on the x-axis according to labels, with 0 indicating "fake" news and 1 "real" news. The quantity of articles belonging to each category is shown on the y-axis. The graph reveals a significant class imbalance, with approximately 15,000 articles labeled as "real" news and only around 4,500 articles labeled as "fake" news.

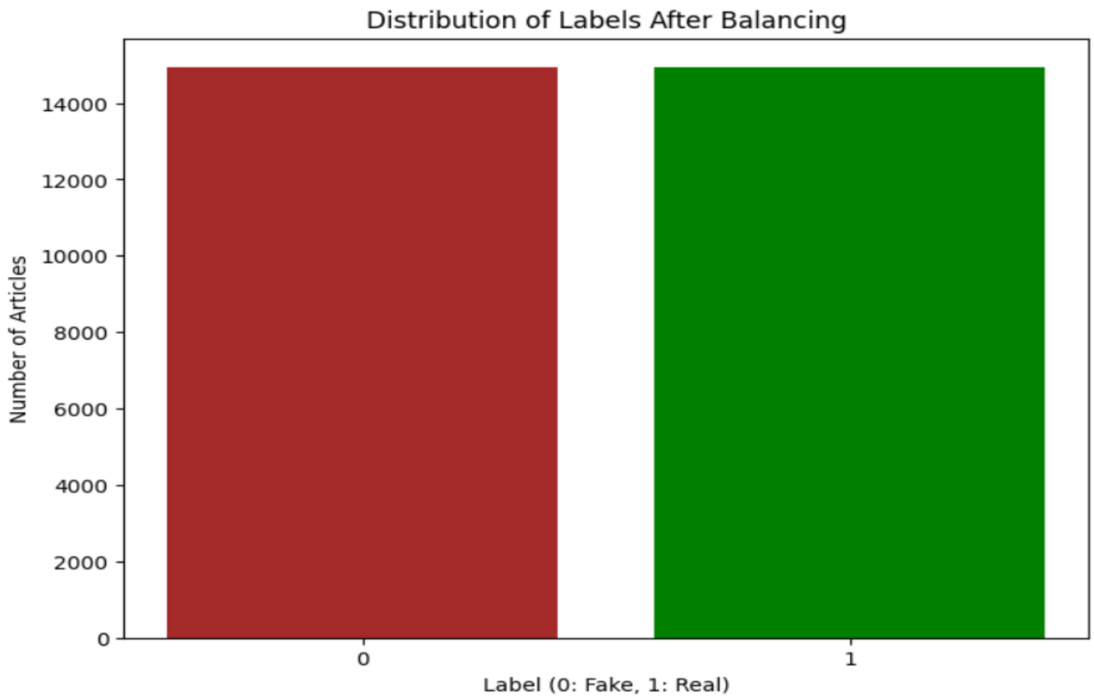

Figure 4.31: Bar Graph for Distribution of Labels After Applying SMOTE Data Balancing on GossipCop Dataset

Figure 4.31 illustrates the distribution of labels after the implementation of SMOTE data balancing to the Gossip Cop dataset. The graphical representation uses bars to present the frequency data between articles recognized as "fake" (0) or "real" (1). Prior to SMOTE operation there existed an unbalanced article distribution featuring fewer "fake" examples compared to "real" examples. The post-SMOTE distribution shows balanced representation between "fake" and "real" articles thus demonstrating that the class imbalance problem received proper resolution through SMOTE.

4.5 Experimental Results

This section shows the experimental analysis of the proposed Roberta sequence classifier model which operates on PolitiFact and Gossip Cop datasets to demonstrate its effectiveness in social media false news classification. The model assessment relied on the confusion matrix Roc curve along with accuracy, precision, recall and f1-scores.

Results of the Proposed Roberta Sequence Classifier on the PolitiFact Dataset

	precision	recall	f1-score	support
0	0.93	0.93	0.93	70
1	0.94	0.94	0.94	85
accuracy			0.94	155
macro avg	0.93	0.93	0.93	155
weighted avg	0.94	0.94	0.94	155

Figure 4.32: Binary Classification Report of Roberta Sequence Classifier on PolitiFact Dataset

The proposed Roberta sequence classifier generates its binary classification report for the PolitiFact dataset as shown in Figure 4.32. The report focusses on the two classes' F1-score, recall, and precision. The precision, recall, and F1-score for class 0 are all 93%, suggesting that the performance is consistent in accurately identifying this class. The precision, recall, and F1-score for class 1 are marginally superior, at 94%. These measures show that the classifier is well-balanced and works effectively for both classes.

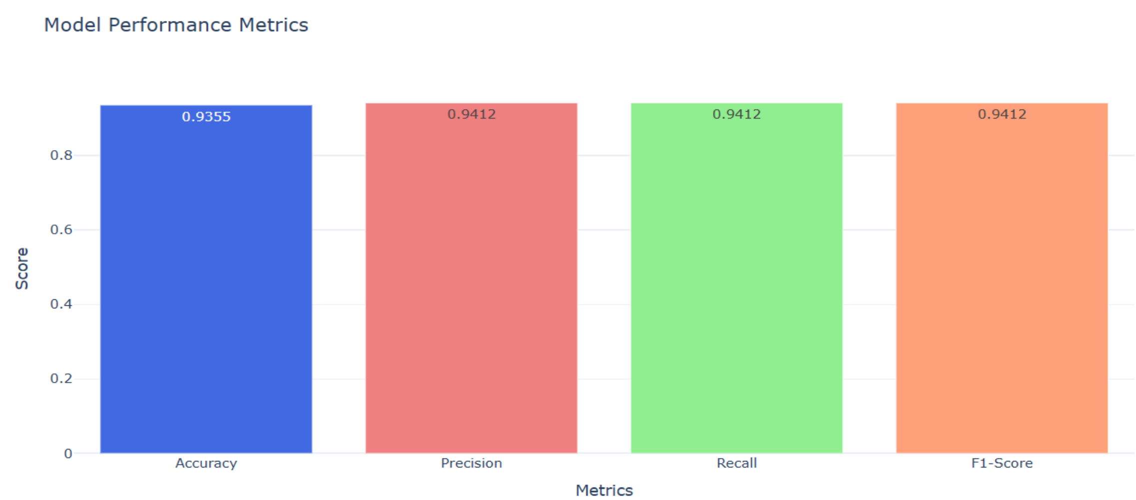

Figure 4.33: Roberta Sequence Classifier’s Performance on PolitiFact Dataset

Figure 4.33 displayed a bar graph representing an efficacy of the Roberta sequence classifier on the PolitiFact dataset in identifying false information inside social media data. To identify the performance, present study utilized commonly used classification metrics, namely accuracy, precision, recall, and f1-score. The bar graph clearly demonstrates that the proposed model achieves a precision of 94.12%, accuracy of 93.55%, re-call of 94.12% and f1-score of 94.12% throughout its testing phase. The results highlight the model's great performance and its effective training.

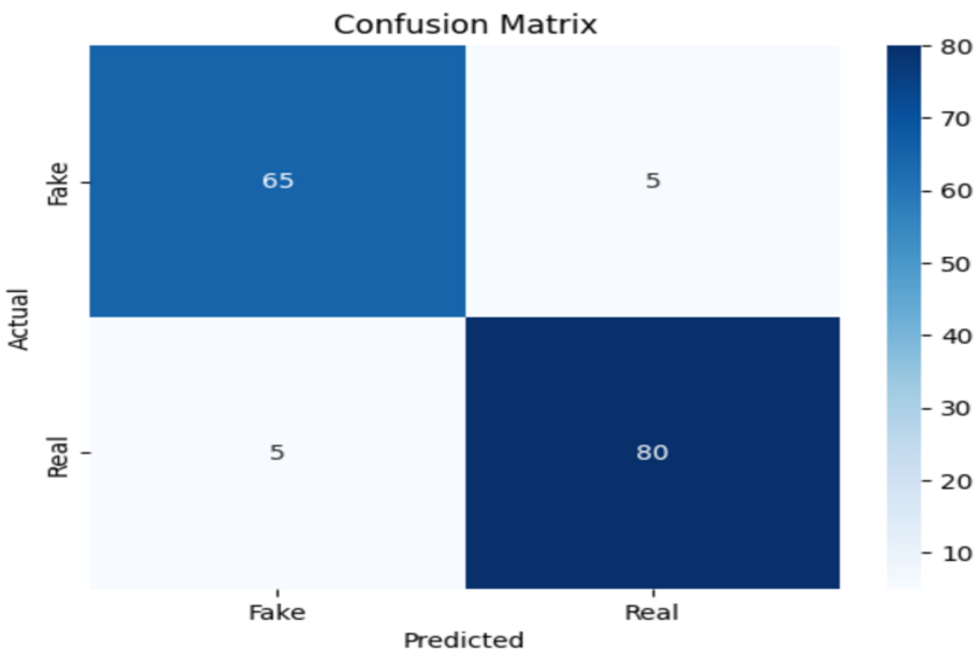

Figure 4.34: Confusion Matrix of Roberta Sequence Classifier on PolitiFact Dataset

Figure 4.34 shows a confusion matrix representing the effectiveness of the Roberta sequence classifier on the PolitiFact dataset in identifying false information inside social media data. The actual label is shown on the y-axis of the matrix, while the anticipated label is shown on the x-axis. Evidence from the confusion matrix demonstrates that the suggested model appropriately labels 65 occurrences as false news and 80 as legitimate news.

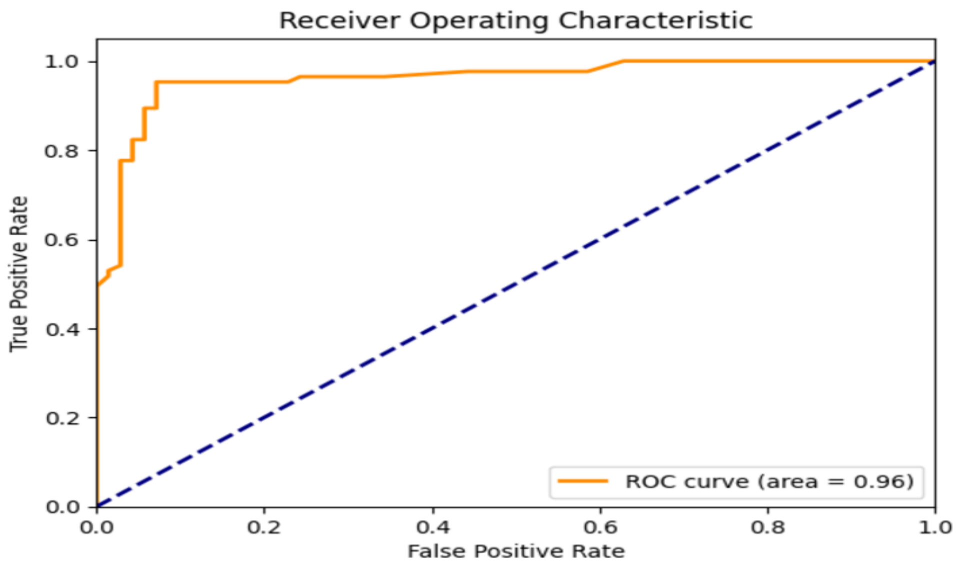

Figure 4.35: ROC Curve of Roberta Sequence Classifier on PolitiFact Dataset

Figure 4.35 displayed a ROC graph representing the efficacy of the Roberta sequence classifier on the PolitiFact dataset in identifying false information inside social media data. The graph's x-axis indicates the FPR, while the y-axis depicts the TPR. The graph clearly demonstrates that the proposed model has a 96% accuracy in classifying fake news.

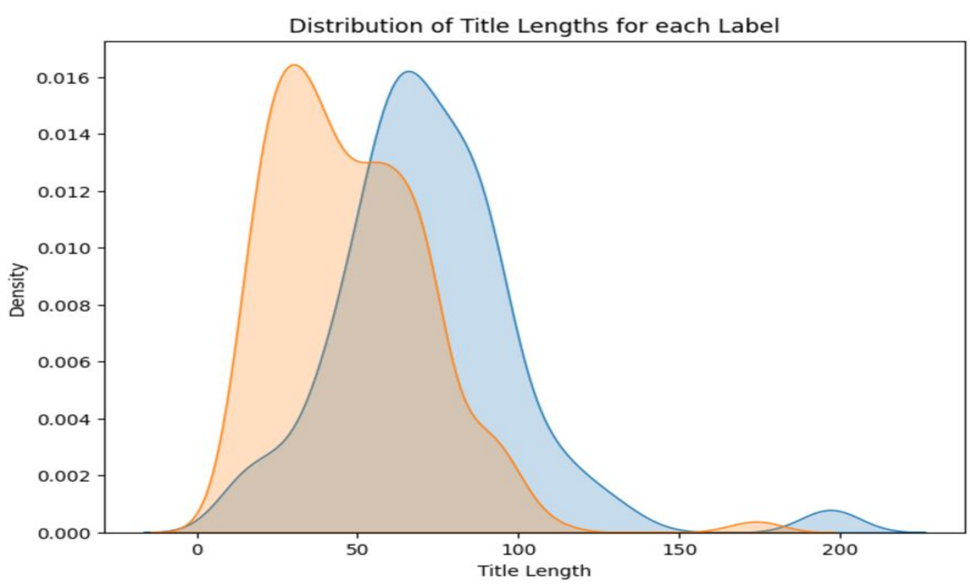

Figure 4.36: Distribution of Title Lengths for Each Label using Roberta Sequence Classifier on PolitiFact Dataset

Figure 4.36 shows the density plots illustrating the distribution of title lengths for each label using the Roberta sequence classifier on the PolitiFact dataset. The x-axis shows the title lengths, ranging from 0 to over 200 characters, while the y-axis represents the density. The two distributions correspond to distinct labels, with the orange curve peaking around 40 characters and the blue curve peaking slightly higher, near 60 characters. Both distributions exhibit a right-skewed pattern, indicating that most titles are relatively short, with only a few extending beyond 100 characters. The density of the orange curve slightly exceeds 0.015 at its peak, whereas the blue curve reaches a similar maximum density slightly further along the x-axis. The overlapping regions suggest some similarity in title length distribution between the two labels, though clear differences in their central tendencies are visible.

Results of the Proposed Roberta Sequence Classifier on the Gossip Cop Dataset

	precision	recall	f1-score	support
0	0.91	0.98	0.94	2951
1	0.98	0.91	0.94	3025
accuracy			0.94	5976
macro avg	0.94	0.94	0.94	5976
weighted avg	0.94	0.94	0.94	5976

Figure 4.37: Binary Classification Report of Roberta Sequence Classifier on Gossip Cop Dataset

Figure 4.37 shows the binary classification report for the proposed Roberta sequence classifier on the Gossip Cop dataset, focusing on precision, recall, and F1-score for the two classes. For class 0, the proposed classifier obtains a precision of 91%, a recall of 98%, and an F1-score of 94%, suggesting a strong capacity to accurately identify instances of this class with a high recall. Similarly, for class 1 model achieves exceptional precision and balanced overall performance, as evidenced by its precision of 98%, recall of 91%, and F1-score of 94%.

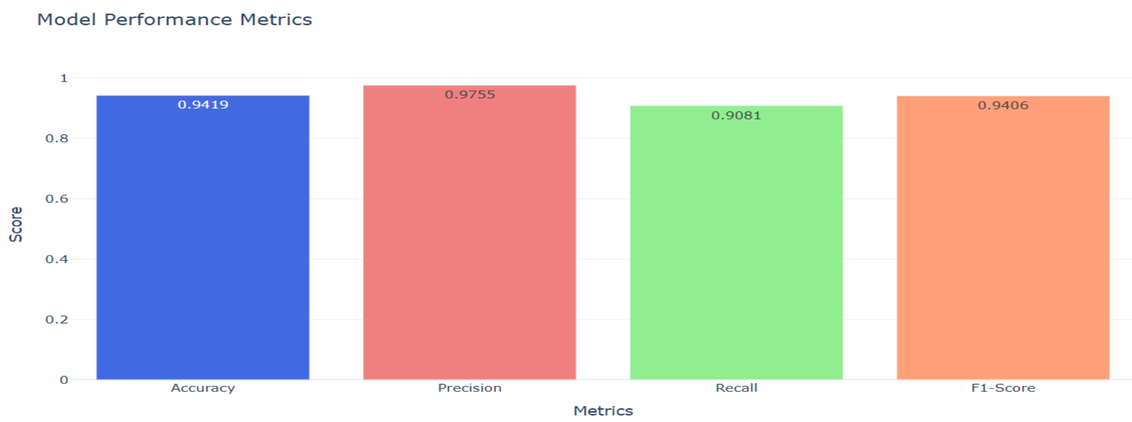

Figure 4.38: Roberta Sequence Classifier's Performance on Gossip Cop Dataset

Figure 4.38 depicts a bar graph representing the efficacy of the Roberta sequence classifier on the Gossip Cop dataset in identifying false information inside social media data. To identify the performance, present study utilized commonly used classification metrics, namely accuracy, precision, recall, and f1-score. The bar graph clearly demonstrates that the proposed model achieves a precision of 97.55%, accuracy of 94.19%, recall of 90.81% and f1-score of 94.06% throughout its testing phase. The results highlight the model's great performance and its effective training.

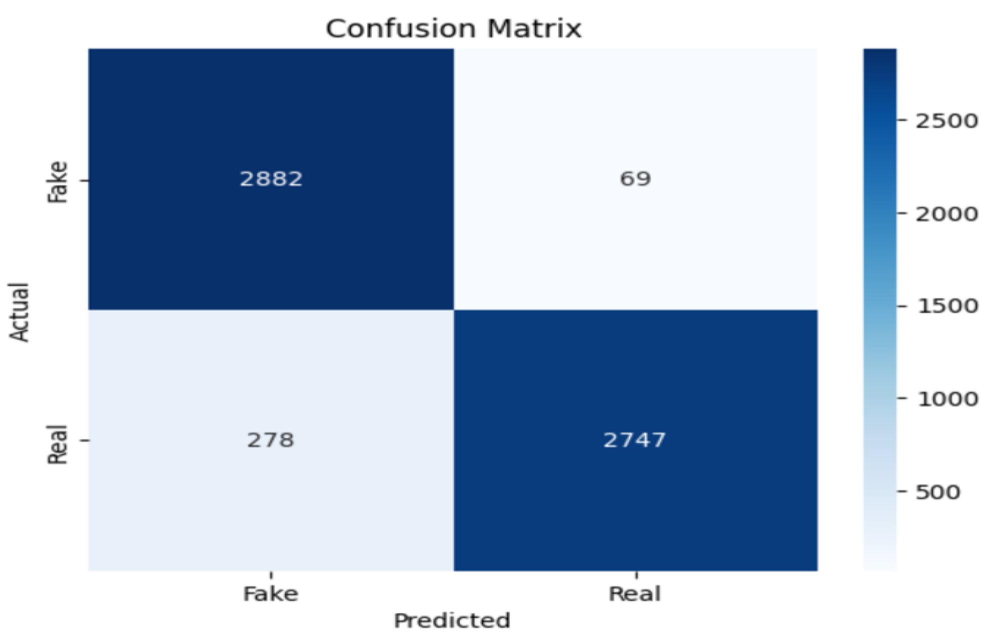

Figure 4.39: Confusion Matrix for Roberta Sequence Classifier on Gossip Cop Dataset

Figure 4.39 shows a confusion matrix representing the effectiveness of the Roberta sequence classifier on the Gossip Cop dataset in identifying false information inside social media data. The actual label is shown on the y-axis of the matrix, while the anticipated label is shown on the x-axis. According to the confusion matrix, the suggested model accurately identifies 28,82 occurrences of false news and 27,47 occurrences of true news.

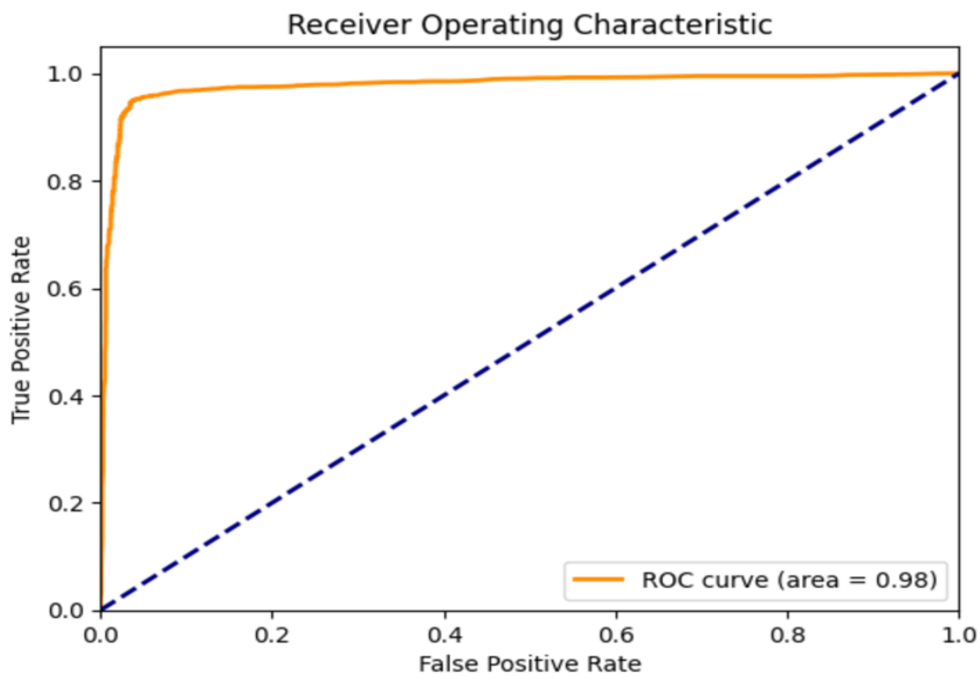

Figure 4.40: ROC Curve of Roberta Sequence Classifier on Gossip Cop Dataset

Figure 4.40 displayed a ROC graph representing the efficacy of the Roberta sequence classifier on the Gossip Cop dataset in identifying false information inside social media data. The graph's x-axis indicates the FPR, while the y-axis depicts the TPR. The graph clearly demonstrates that the proposed model has a 98% accuracy in classifying fake news.

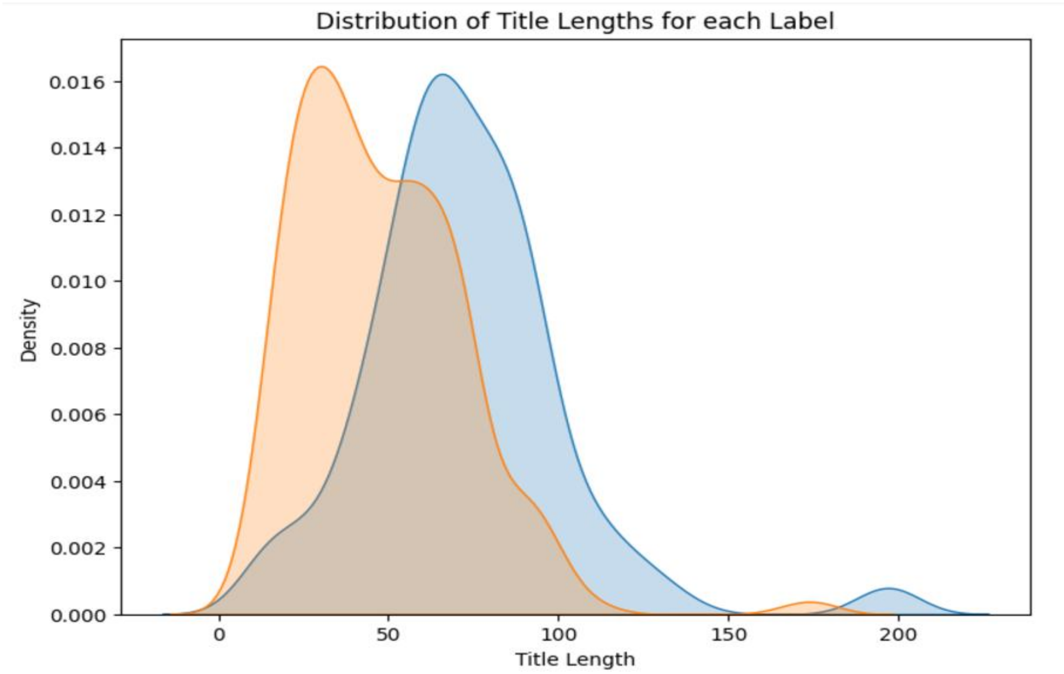

Figure 4.41: Distribution of Title Lengths for Each Label using Roberta Sequence Classifier on Gossip Cop Dataset

Figure 4.41 displays the distribution of title lengths for both "Fake" and "Real" news articles in the Gossip Cop dataset, as identified by the Roberta sequence classifier. The x-axis of graph indicates the title length ranging from 0 to 200 characters, while the y-axis represents the density of article with a specific title length. In the figure, blue curve and shaded area show the distribution of "Real" news while the orange curve and shade area depict the distribution of "Fake" news. The calculated distribution for both 'Fake' and 'Real' news titles is almost the same, with most of the titles having from 0 to 100 characters. However, there is a slight difference when it comes to the density of the news titles with "Real" news title having a slightly higher density in the 50 to 100-character range than "Fake" news titles.

CHAPTER V:

ANALYSIS AND DISCUSSION

5.1 Comparative Analysis

This section provides a comparative examination of the outcomes that have been achieved via the construction of a system to detect false news using the deep learning model.

Comparison of the Proposed DL Models on Two Datasets

In this section, compare the suggested Roberta sequence classifier model to two datasets and identify which one is more accurate in identifying the fake news.

Table 5.1: Comparison of Proposed DL Model on Two Different Datasets

Parameters	**PolitiFact Dataset**	**Gossip Cop Dataset**
	Roberta Sequence Classifier	**Roberta Sequence Classifier**
Accuracy	93.55	94.19
Precision	94.12	97.55
Recall	94.12	90.81
F1-Score	94.12	94.06

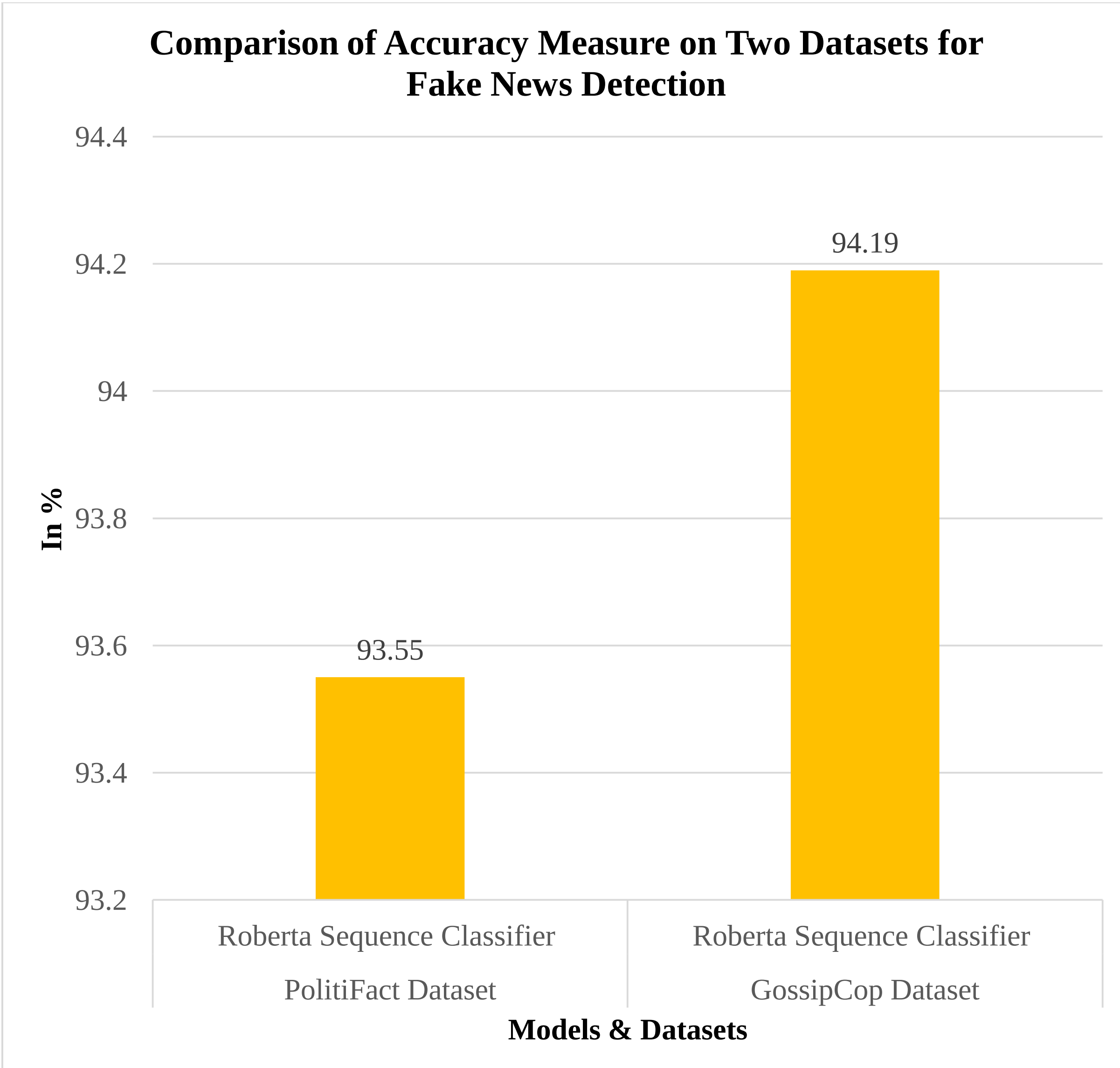

Figure 5.1: Bar Graph for Comparison of Accuracy of Roberta Sequence Classifier on Two Different Datasets

Figure 5.1 shows the accuracy measurement of the proposed sequence classifier Roberta which was tested on two separate sets of data. The cross-validation is to evaluate the capability of proposed model in identifying fake news on two datasets in order to decide which of the two datasets best suits the current study. From the figure, the accuracy graph shows a very high performance of the proposed model on Gossip Cop, with an accuracy of 94.19%. However, the kind of model also performs better in the case of the PolitiFact dataset with an accuracy of 93.55%, slightly lower than that of the Gossip Cop dataset. This proposed Roberta model's improved accuracy levels are also demonstrated when applied to data from the Gossip Cop dataset.

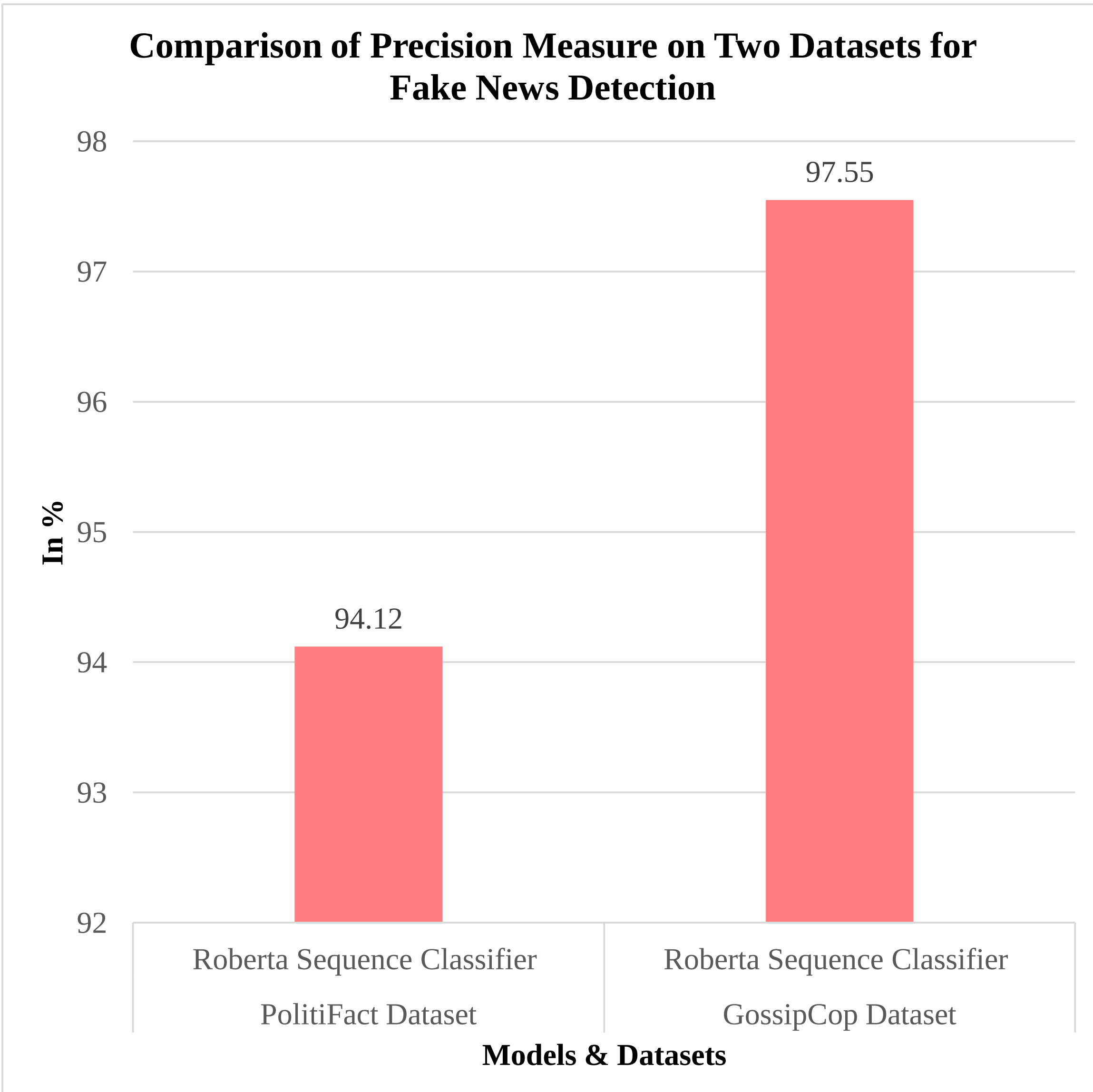

Figure 5.2: Bar Graph for Comparison of Precision of Roberta Sequence Classifier on Two Different Datasets

Figure 5.2 shows the precision rates of the proposed Roberta sequence classifier when used with two different datasets. The criteria for the identification of false news performance evaluation in the proposed model is based on the comparison of two different datasets for the selection of dataset for this study. The precision graph in the figure also points clearly to the fact that among all the systems, the proposed model has the highest precision of 97.55% on Gossip Cop. On the other hand, the suggested model has a higher accuracy of 99.77% on the PolitiFact dataset while having slightly less precision of 94.12% than the Gossip Cop dataset. The proposed Roberta model yields higher precision rates as compared to the other models in analyzing the Gossip Cop dataset.

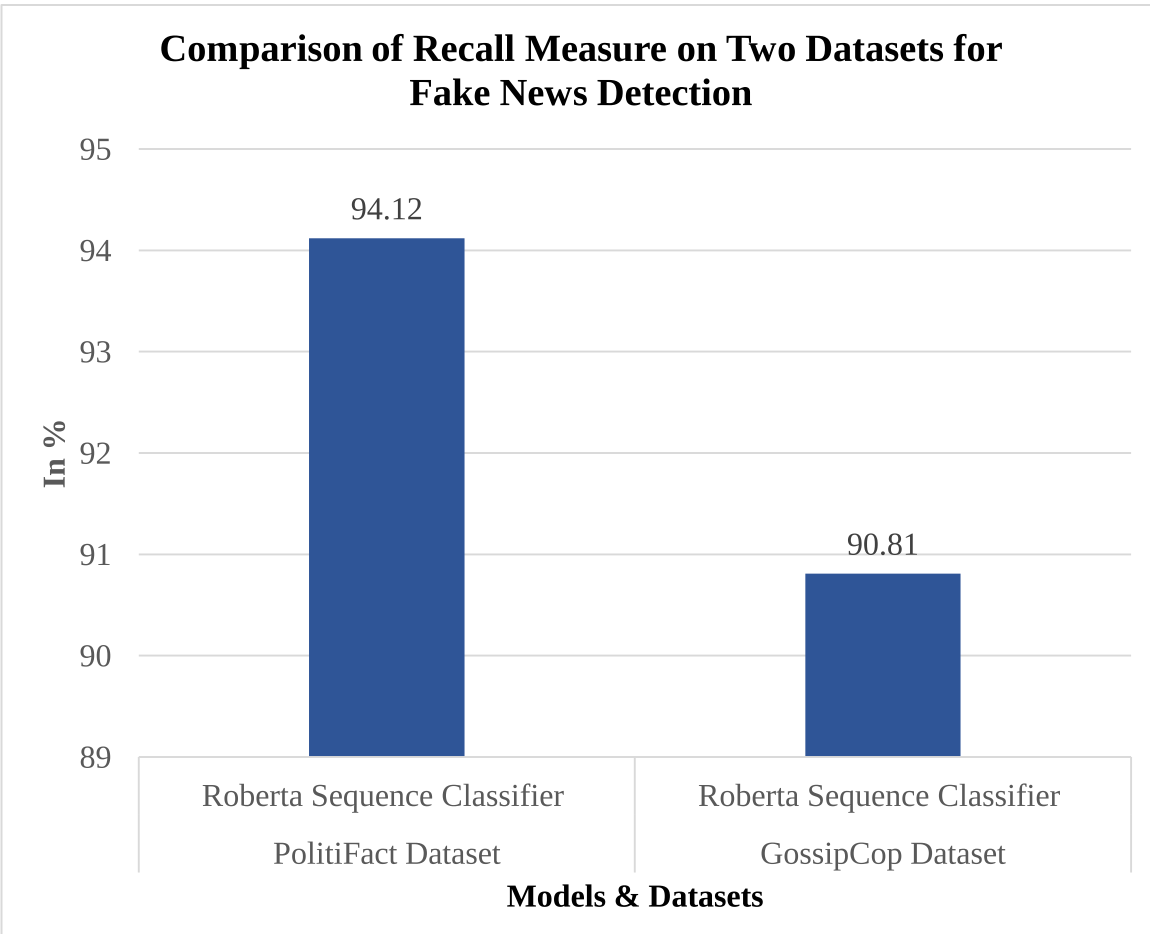

Figure 5.3: Bar Graph for Comparison of Recall of Roberta Sequence Classifier on Two Different Datasets

Figure 5.3 shows the performance of the proposed Roberta sequence classifier as concerns the recall measurement across the datasets. It is used to establish the capability of the proposed model in identifying fake news on two different datasets and to decide which of the datasets is appropriate for the current study. As reflected in the figure, the recall graph shows the highest result of the proposed model on PolitiFact with the highest recall of 94.12%. However, the proposed model performs even better on the accuracy measure on the Gossip Cop dataset with a recall of 90.81 % as compared to the PolitiFact dataset. The proposed Roberta model is notably better in the aspect of recall when compared to other models for fact-checking on the PolitiFact dataset.

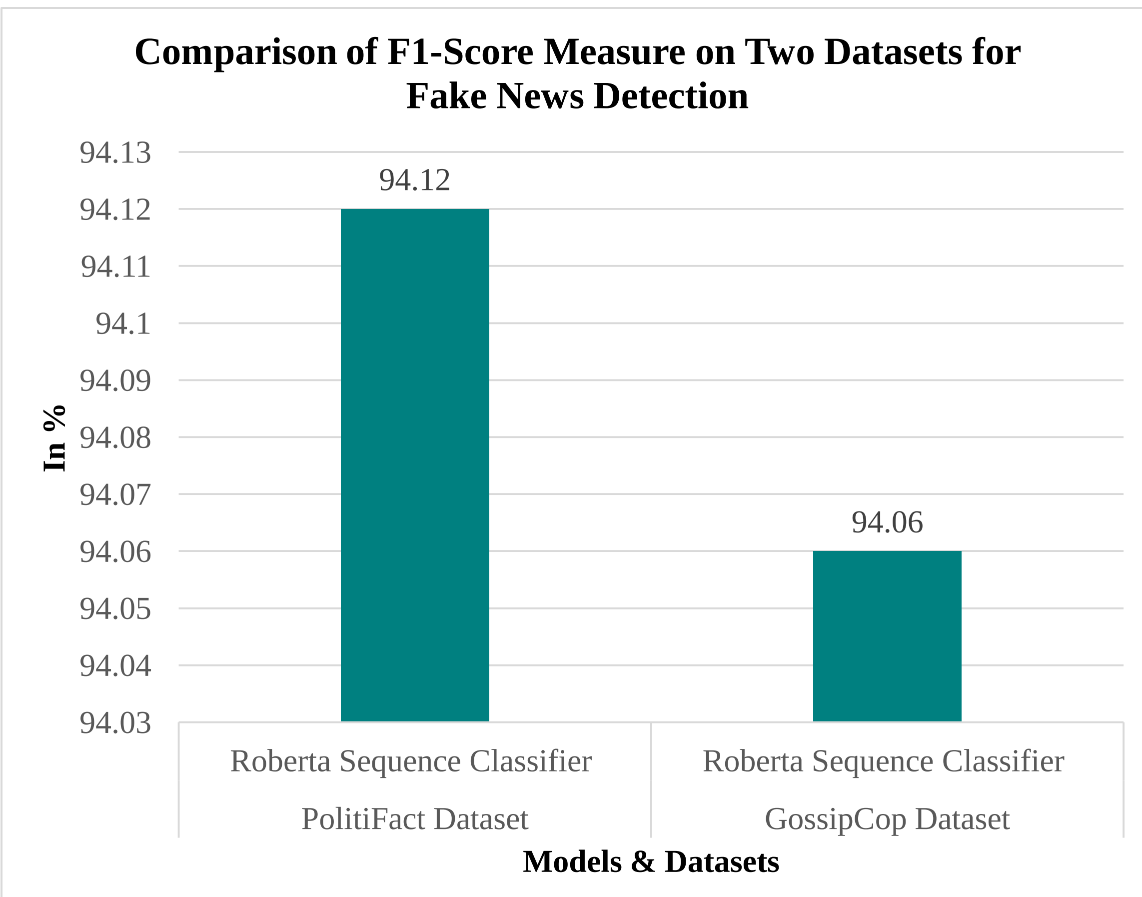

Figure 5.4: Bar Graph for Comparison of F1-Score of Roberta Sequence Classifier on Two Different Datasets

The f1-scores obtained by the proposed Roberta sequence classifier on two different datasets as mentioned in the comparison are depicted in figure 5.4. The proposed model is evaluated in this study with the aim of comparing the results of false news identification with other datasets in other studies in order to establish which of the dataset best suits the current study. From the f1-score graph in the figure, it is evident that the proposed model yields highly accurate results for PolitiFact with a f1-score of 94.12%. On the other hand, the proposed model showed higher accuracy and achieved slightly lower f1-score of 94.06% on the Gossip Cop dataset compared to the PolitiFact dataset. Overall, using the Roberta model as suggested for the PolitiFact dataset in this comparison resulted in higher f1-score.

Comparison of Base and Proposed Models on PolitiFact Dataset

Table 5.2: Comparison of Base and Proposed Model on PolitiFact Dataset for Fake News Detection

Parameters	Base Model			Proposed Model
	TChecker(GabAllah, Sharara and Rafea, 2023)	dEFEND(Shu *et al.*, 2019)	TCNN-URG(Qian *et al.*, 2018)	Roberta Sequence Classifier
Accuracy	90.25	90.00	71.20	**93.55**
Precision	92.75	90.00	72.30	**94.12**
Recall	86.48	90.00	71.20	**94.12**
F1-Score	89.51	90.00	72.20	**94.12**

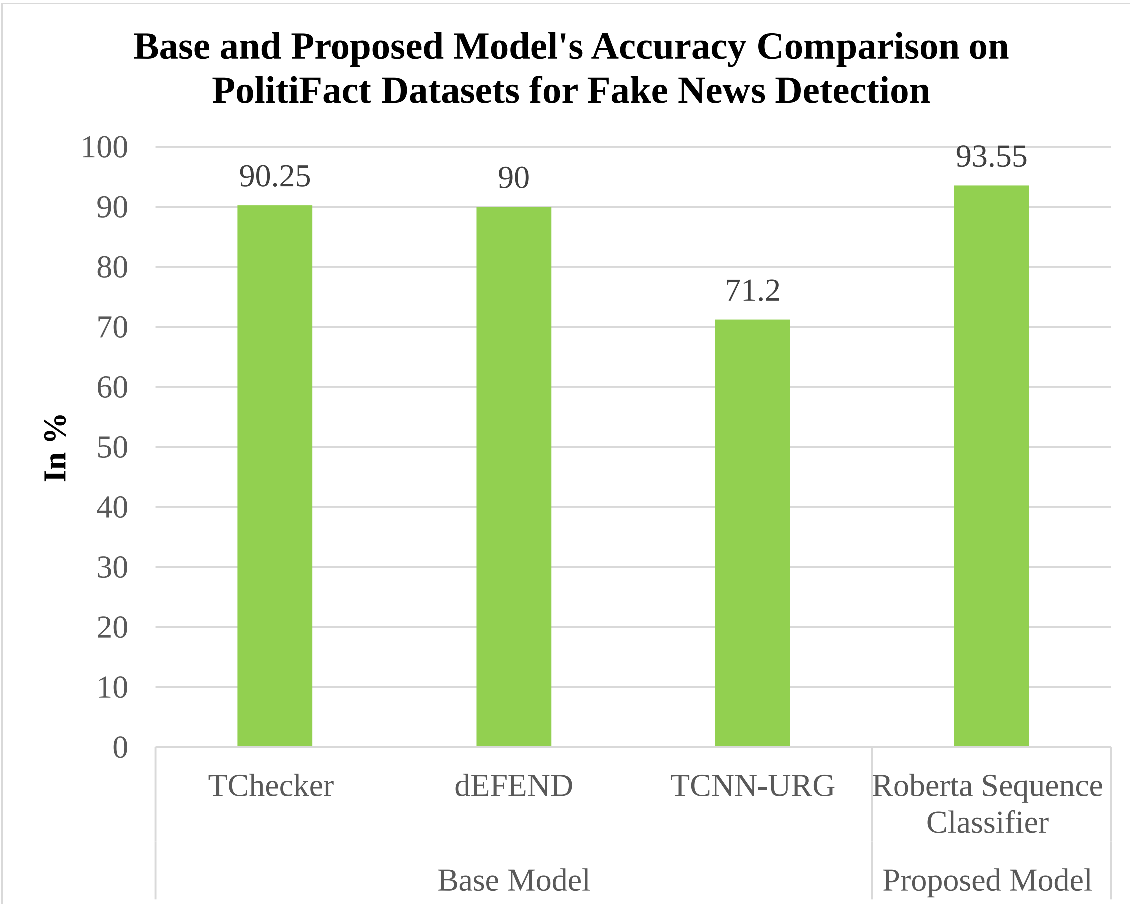

Figure 5.5: Comparison of Accuracy Between Base and Proposed Model on PolitiFact Dataset for Fake News Detection

The above-mentioned Table 5.2 and Figure 5.5 presents a bar graph comparing the accuracy of base models (TChecker, dEFEND, TCNN-URG) and a proposed Roberta sequence classifier model in detecting fake news on the PolitiFact dataset. In the figure, the graph clearly demonstrates that the base model TChecker has an accuracy of 90.25, dEFEND has an accuracy of 90%, and TCNN-URG has an accuracy of 71.2%, while the proposed model has an accuracy of 93.55%. The proposed model proves superior to the base model at detecting fake news on the PolitiFact dataset according to the accuracy results.

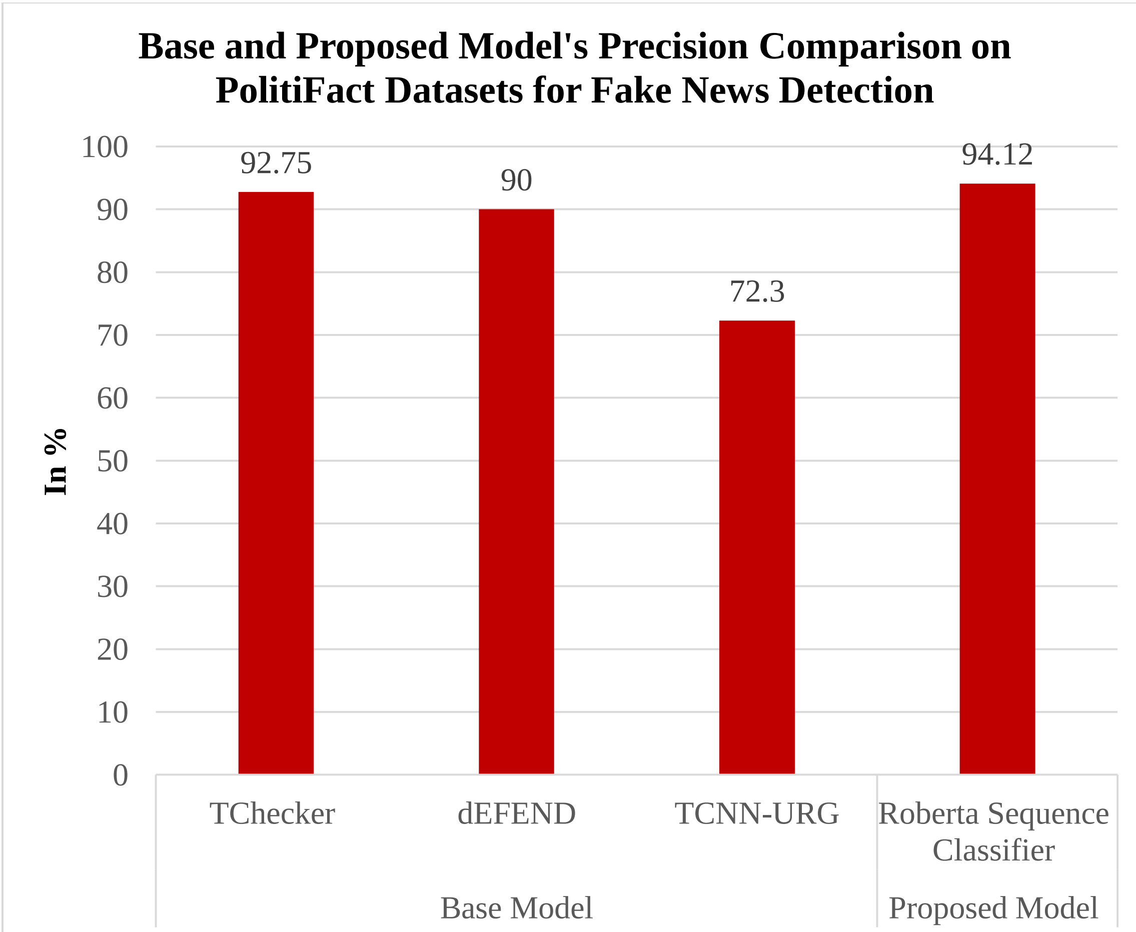

Figure 5.6: Comparison of Precision Between Base and Proposed Model on PolitiFact Dataset for Fake News Detection

The above-mentioned Table 5.2 and Figure 5.6 present a bar graph comparing the precision of base models (TChecker, dEFEND, TCNN-URG) and a proposed Roberta sequence classifier model in detecting fake news on the PolitiFact dataset. From the figure it can also be noted that the proposed model, TChecker, dEFEND and TCNN-URG model, respectively have a precision of 92.75%, 90%, 72.30%, and 94.12% respectively. The results obtained above reveal that the proposed model is better than the baseline when it comes to detecting false information within the PolitiFact dataset.

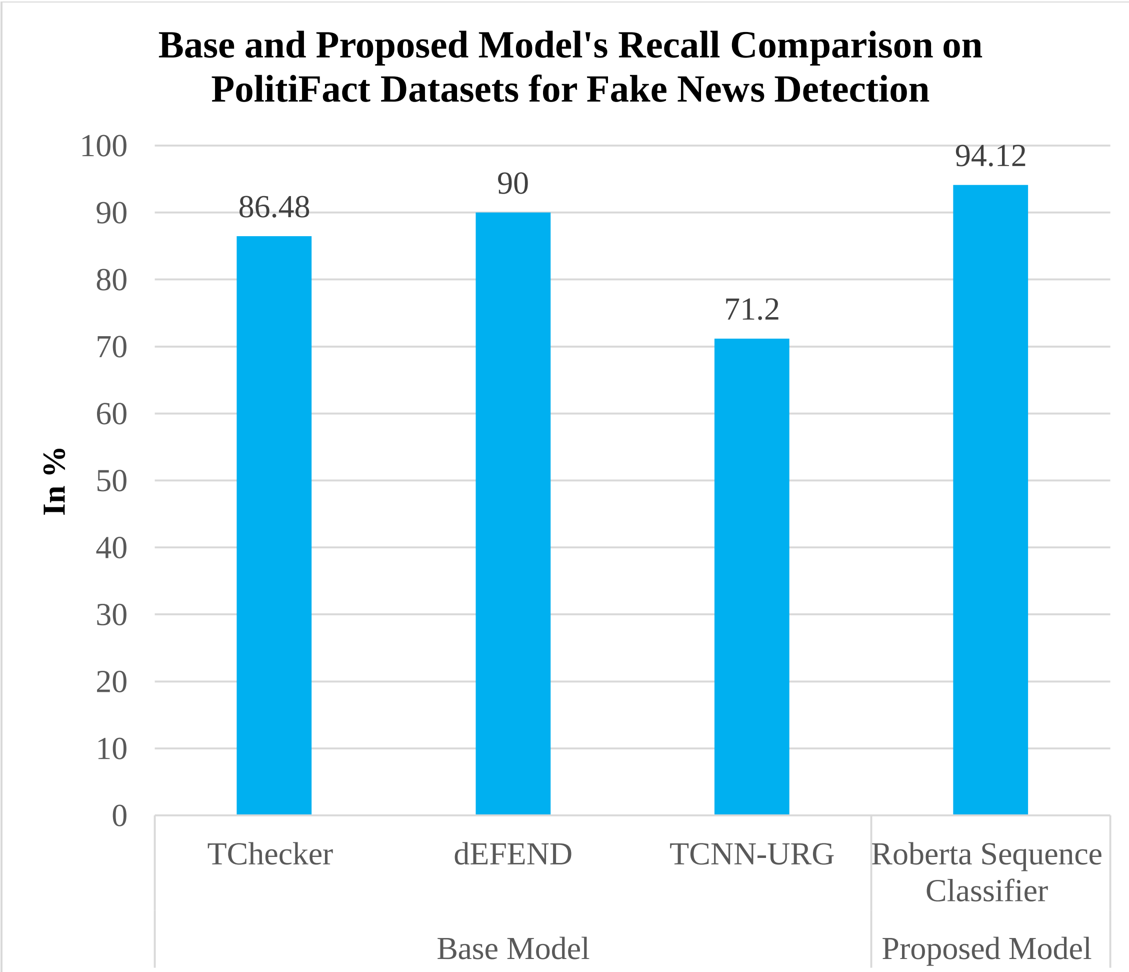

Figure 5.7: Comparison of Recall Between Base and Proposed Model on PolitiFact Dataset for Fake News Detection

Table 5.2 and Figure 5.7 above shows the bar graph that captures the recall of the baselines (TChecker, dEFEND, TCNN-URG) and the proposed RoBERTa sequence classifier model on the PolitiFact dataset. This graph clearly shows that the base model TChecker has a recall of 86.48%, dEFEND model has a recall of 90%, TCNN-URG model has a recall of 71.2% and the proposed model that has a recall of 94.12%. These scores clearly show that the proposed model is better than the base model in terms of recall when it comes to Fake News detection in the PolitiFact set.

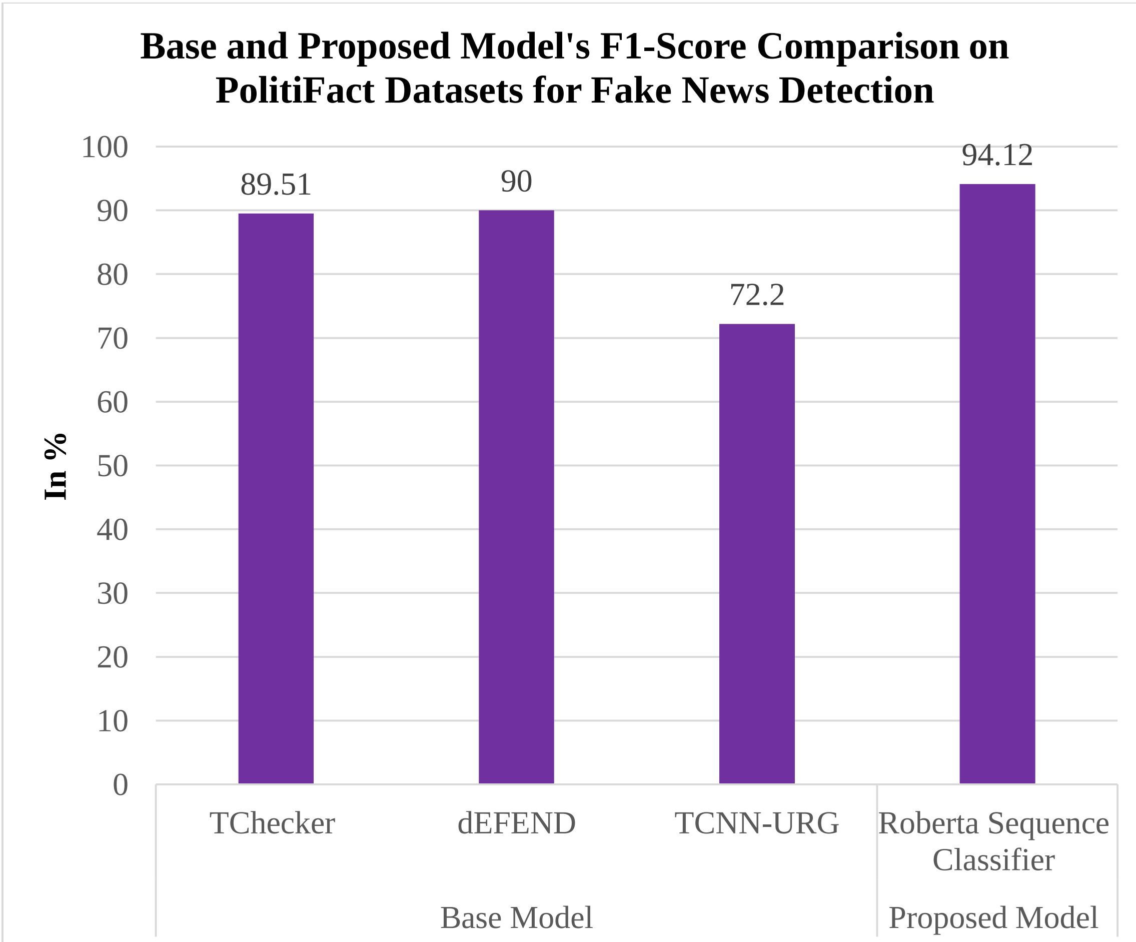

Figure 5.8: Comparison of F1-Score Between Base and Proposed Model on PolitiFact Dataset for Fake News Detection

The above-mentioned Table 5.2 and Figure 5.8 present a bar graph comparing the f1-score of base models (TChecker, dEFEND, TCNN-URG) and a proposed Roberta sequence classifier model in detecting fake news on the PolitiFact dataset. In the figure, the graph clearly demonstrates that the base model TChecker has a f1-score of 89.51%, dEFEND has a f1-score of 90%, and TCNN-URG has a f1-score of 72.20%, while the proposed model has a f1-score of 94.12%. This comparison demonstrates that when it comes to recognizing false news in the PolitiFact dataset, the suggested model outperforms the baseline.

Comparison of Base and Proposed Models on Gossip Cop Dataset

Table 5.3: Comparison of Base and Proposed Model on Gossip Cop Dataset for Fake News Detection

Parameters	Base Model			Proposed Model
	TChecker(GabAllah, Sharara and Rafea, 2023)	**dEFEND** (Shu *et al.*, 2019)	**TCNN-URG** (Qian *et al.*, 2018)	**Roberta Sequence Classifier**
Accuracy	89.50	80.80	73.90	**94.19**
Precision	89.20	80.80	71.50	**97.55**
Recall	89.71	80.80	52.10	**90.81**
F1-Score	89.46	80.80	60.30	**94.06**

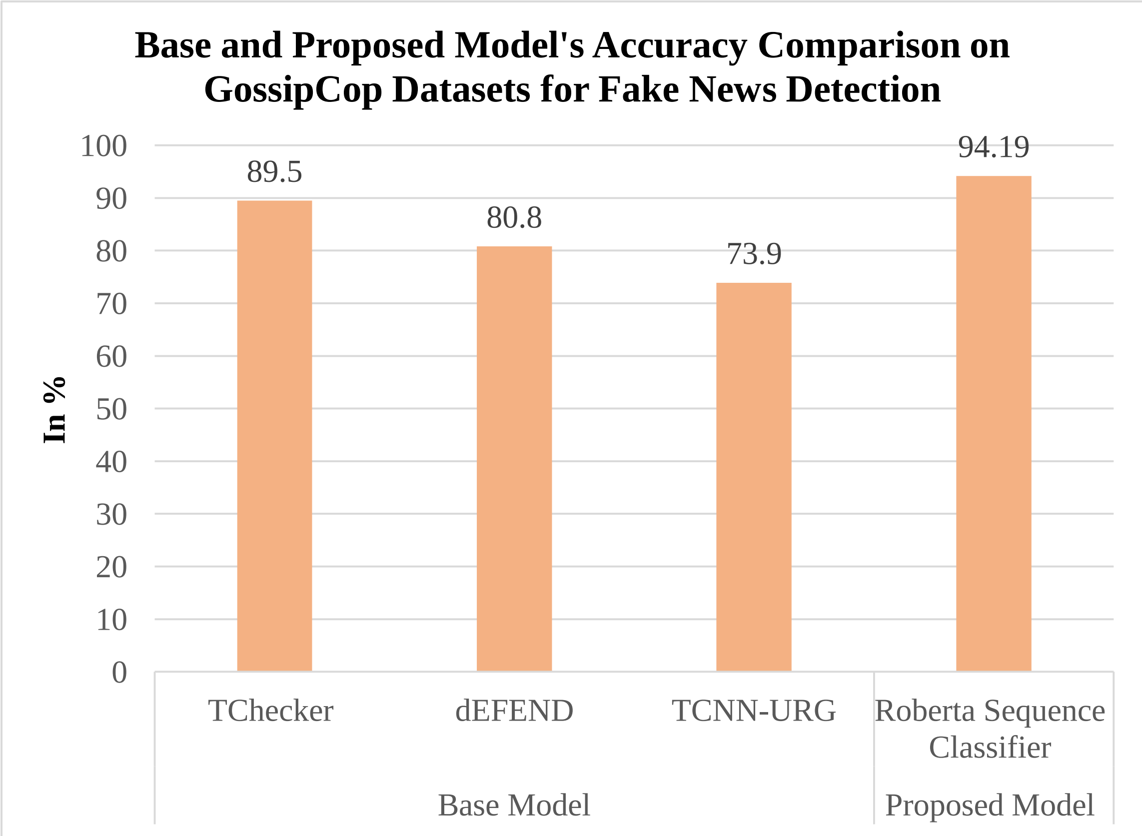

Figure 5.9: Comparison of Accuracy Between Base and Proposed Model on Gossip Cop Dataset for Fake News Detection

The accuracy of each base model and the proposed model for the fake news detection dataset known as Gossip Cop is demonstrated in Figure 5.9. Based on the base models, TChecker has the most diverse accuracy of 89.5%, whereas dEFEND recognizes 80.8%, and TCNN-URG recognizes 73.9%. For the Roberta Sequence classifier, the final accuracy score of 94.19% beats all the base models with a significant difference. This shows a very promising result that supports the notion that the proposed model could have a high level of success in identifying fake news. The chart further emphasizes the disparity in the performance where it shows that the current base models are inadequate in dealing with large-scale datasets like Gossip Cop. These improvements in the architecture of the Roberta Sequence Classifier and its capability to capture more context make the model more dependable for fake news detection than the base models.

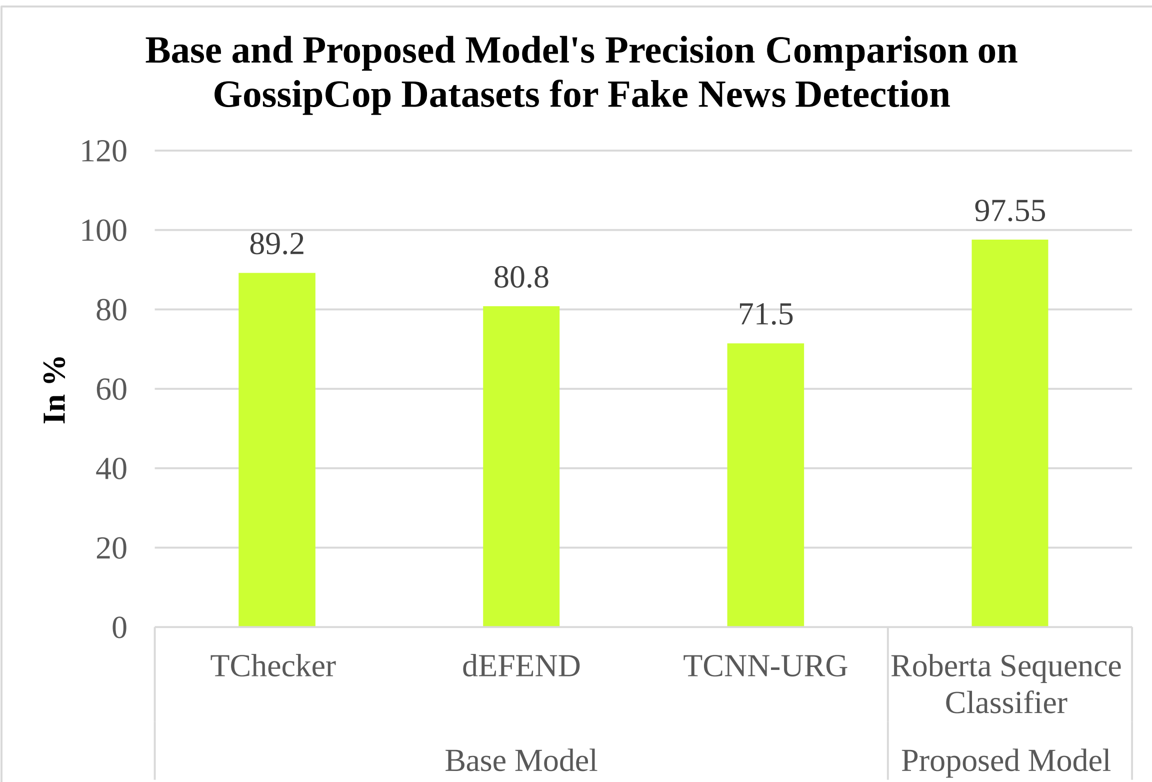

Figure 5.10: Comparison of Precision Between Base and Proposed Model on Gossip Cop Dataset for Fake News Detection

Figure 5.10 illustrates a bar graph that compares the precision of base models (TChecker, dEFEND, TCNN-URG) with a suggested Roberta sequence classifier model in identifying fake news on the Gossip Cop dataset. The x-axis classifies the models into base and suggested categories, whilst the y-axis denotes accuracy expressed as a percentage. The graph indicates that the suggested Roberta sequence classifier attains the maximum accuracy at 97.55%, markedly surpassing the baseline models. TChecker and dEFEND have comparable precisions of around 89.2% and 80.8%, respectively, whilst TCNN-URG has the lowest accuracy at 71.5%. The proposed model proves superior than the baseline for detecting fake news when applied to the Gossip Cop dataset.

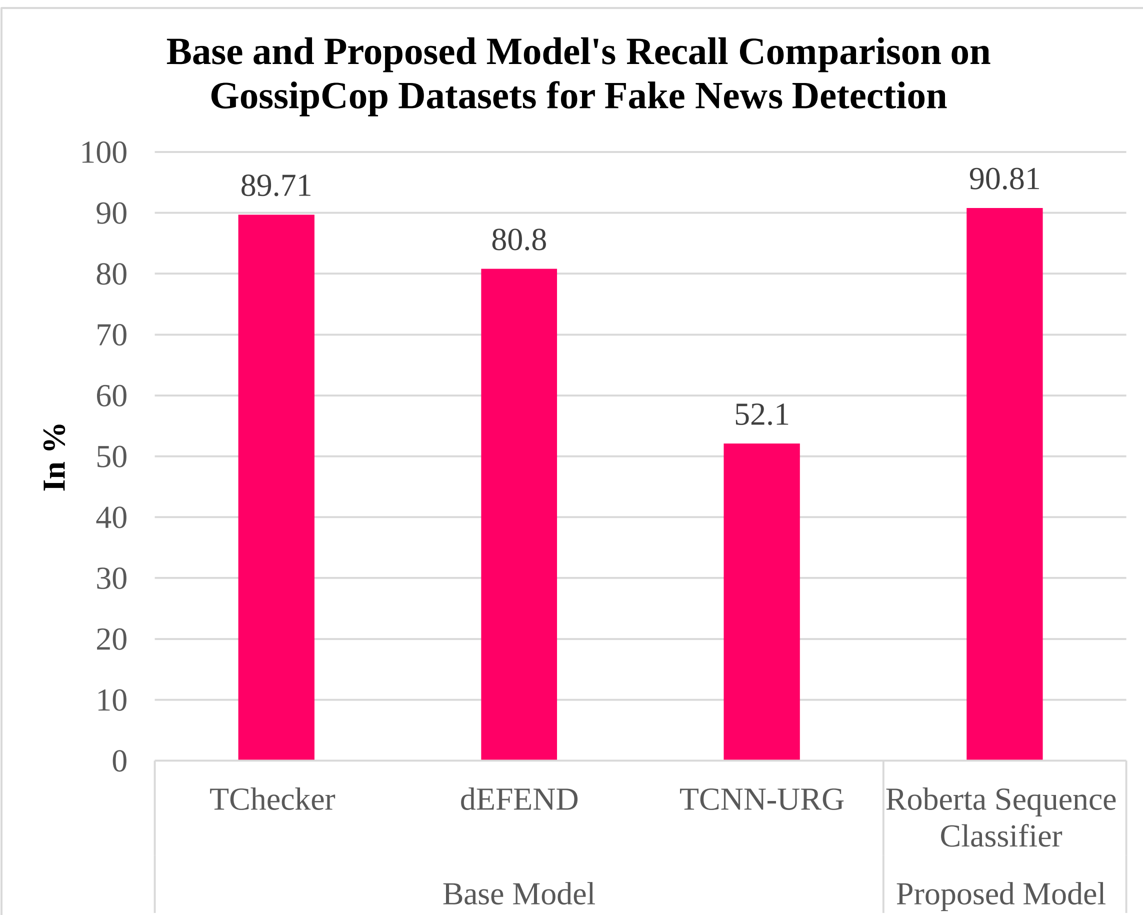

Figure 5.11: Comparison of Recall Between Base and Proposed Model on Gossip Cop Dataset for Fake News Detection

The Gossip Cop dataset underwent recall benchmarking using TChecker, dEFEND, TCNN-URG base models and a sequence classifier built from Roberta as shown in Figure 5.11. The base models can be found horizontally on the x-axis with proposal models represented vertically on the y-axis of the graph and plotted in percentage recall. Through 90.81% recall, the proposed sequence classifier that is built using Roberta supersedes the achievement of base models. TChecker together with dEFEND display comparable recall percentages which stand at 89.71% and 80.8%. TCNN-URG demonstrates the lowest recall at 52.1%. The suggested model demonstrates its effectiveness at identifying genuine fake news occurrences in the Gossip Cop dataset according to these evaluation results.

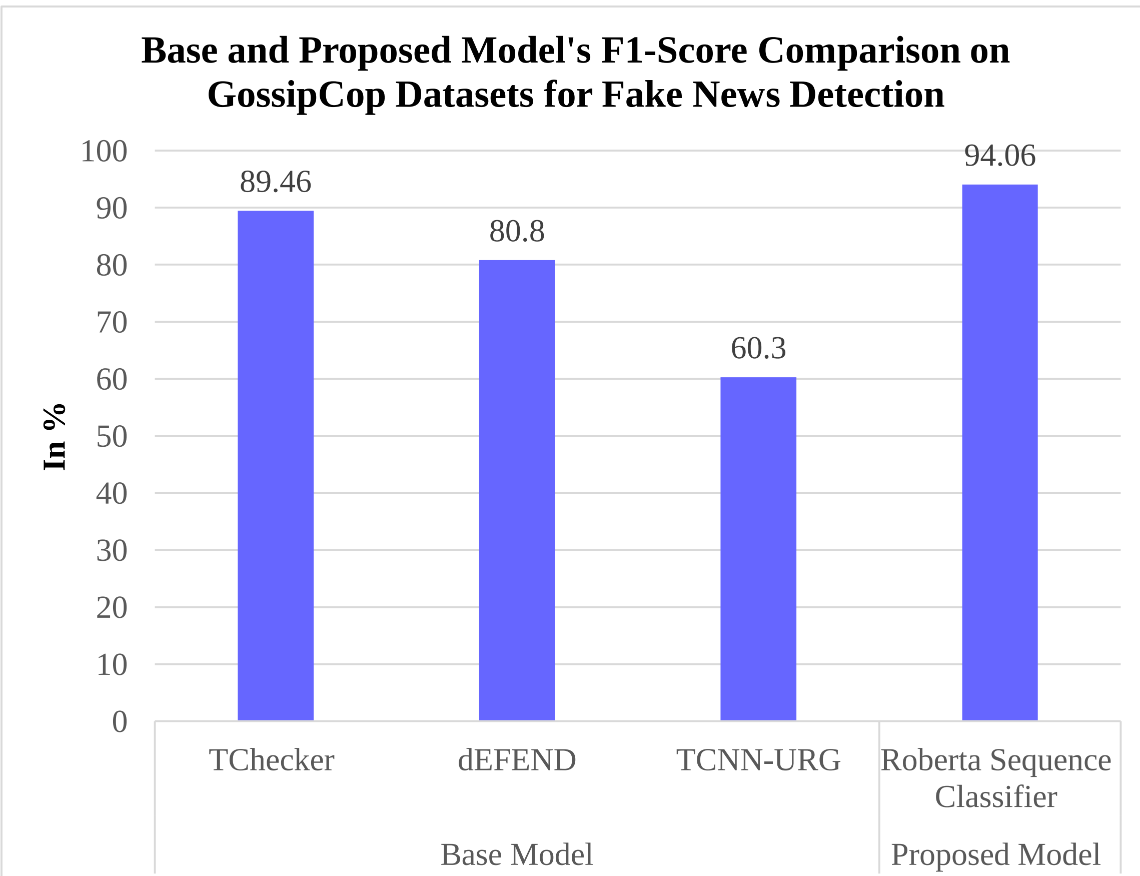

Figure 5.12: Comparison of F1-Score Between Base and Proposed Model on Gossip Cop Dataset for Fake News Detection

Figure 5.12 shows the F1-score assessment between the base models TChecker, dEFEND, and TCNN-URG and the proposed Roberta sequence classifier which analyzes fake news detection on the Gossip Cop dataset. The x-axis categorizes the models into base and proposed models, while the y-axis represents F1-score in percentage. Results demonstrate that the proposed method achieves superior performance by obtaining an F1-score measurement of 94.06%. The F1-score results from TChecker and dEFEND match at around 89.46% and 80.8% respectively. TCNN-URG demonstrates the lowest F1-score at 60.3%. Results prove the proposed model excels above the baseline regarding Gossip Cop fake news detection capability.

5.2 Comparison with Human Fact-Checkers

The model's predictions were compared with expert fact-checker decisions on a subset of the dataset. Roberta achieved a high agreement rate with human annotators. However, it fell short in nuanced and context-heavy scenarios that required up-to-date real-world knowledge, sarcasm detection, or complex narrative interpretation. Expert fact-checkers, equipped with current events awareness and domain-specific insight, were more accurate in these instances. This analysis suggests that a hybrid human-AI framework could significantly enhance the accuracy and scalability of fake news detection systems. In this approach, Roberta can be employed as a rapid screening tool to flag potentially false content, which is then reviewed by trained human experts. This division of labor allows the system to leverage AI's processing speed and consistency while also drawing on human interpretive capabilities for complex cases. Such synergy could lead to more trustworthy and efficient misinformation management across platforms.

5.3 Discussion

Reliable detection of fake news through algorithms represents a critical issue for the current digital age because these algorithms need to identify misguiding content within diverse datasets correctly. To determine how well the suggested Roberta sequence classifier model addresses this issue, it has been tested on two popular datasets, PolitiFact and Gossip Cop. The findings show that, but with minor variances, the model works very well on both datasets, attaining excellent accuracy, precision, recall, and F1-scores. The results of the evaluation of Gossip Cop showed the accuracy of 94.19% and precision of 97.55%, so the opportunities of working with large scale data in variations though with a small number of incorrect results. High recognition of fake news is evident in its performance on PolitiFact that has a 94.12% recall alongside its 94.12 %F1-score supposed to the format of the dataset. The proposed model outperforms TChecker, dEFEND, and TCNN-URG since it comprehensively meets the needs of both datasets by considering various contexts and the model architecture comprehensively. These results suggest that even though there is an improvement with the proposed model on various datasets, Gossip Cop is the more appropriate dataset for this research since the model has the best results overall to determine false news in emergent large-scale scenarios.

5.4 Advantages of the Proposed Fake News Detection Methodology

1) High Accuracy and Robust Performance

- The Roberta sequence classifier achieves 93.55% accuracy on PolitiFact and 94.19% accuracy on Gossip Cop, outperforming existing models such as TChecker, defend, and TCNN-URG.
- High precision and recall values ensure that the model effectively identifies fake news while minimizing false positives and false negatives.

2) Effective Handling of Class Imbalance

- The use of random oversampling balances the dataset, preventing bias towards the majority class and improving model generalization.

3) Comprehensive Pre-processing Pipeline

- Data cleaning, normalization, tokenization, and noise removal ensure high-quality input, enhancing the effectiveness of text representation.

4) State-of-the-Art NLP Model

- Roberta's context-aware deep learning approach enables better text understanding compared to traditional machine learning models or earlier transformer architectures.

5) Scalability and Adaptability

- The methodology is scalable and can be adapted to other languages and datasets with fine-tuning, making it suitable for diverse social media platforms.

6) Exploratory Data Analysis for Insight Generation

- EDA uncovers hidden patterns and trends in misinformation, aiding further research in fake news detection.

7) Robust Evaluation Metrics

- The study employs accuracy, precision, recall, F1-score, confusion matrix, and AUC, ensuring a comprehensive assessment of model performance.

8) Potential for Real-Time Applications

- With optimization (e.g., Distil Roberta or ALBERT for efficiency), the model can be adapted for real-time detection on social media platforms.

9) Outperforms Traditional and Deep Learning Baselines

- Compared to prior models, Roberta-based classification demonstrates superior performance in detecting misleading information.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

These days, everyone gets their news from a multitude of websites. For newsletters and newsreaders alike, it means going digital, which is more convenient and faster than the old ways of getting information. Social media has made it simpler to communicate information, which is why news is rapidly spreading among millions of users on sites like Facebook, Twitter, and others. There may be misleading news on these social media sites since it is so simple to create them. The potential for fake news to destabilize governments and contemporary society has made it a top priority. Victims and communities alike can suffer greatly from the propagation of misinformation. It is crucial to identify incorrect information early on in order to put a halt to its spread and protect innocent individuals from fake news broadcasts. Some of the most efficient and successful ways to identify hoaxes are those involving natural language processing, although there are others. This paper discusses the use of deep learning (DL) and natural language processing (NLP) in fake news detection and seeks to identify areas that need improvement to increase the models' effectiveness. The study utilizes two datasets from PolitiFact and Gossip Cop and the Roberta sequence classifier, which is based on BERT, as an upgraded version and applies dynamic masking, increases a sequence length, and uses a more extensive vocabulary. Preprocessing steps used include tokenization, data cleaning, and oversampling for handling issues to do with class imbalance. This study shows that the Roberta model achieved high scores of accuracies, precision, recall, and F1-score, which are more than 93% and significantly outcompeted traditional models of TChecker, dEFEND, and TCNN-URG. As for the comparisons by the dataset, it obtains a very high precision of 97.55% on the Gossip Cop dataset while having a relatively higher recall of 94.12% compared with other models on the PolitiFact dataset, which proves that it can perceive subtle contextual and linguistic features. Such results indicate the model's applicability in reducing the effects of misinformation industries like media, cybersecurity, and public relations, where timely and accurate fake news identification is crucial. Moreover, the current work combines the benefits of DL and NLP, offers a solid foundation for fighting fake news and suggests practical scenarios while sharing valuable insights into scholarly innovations.

6.2 Implications

The impact of this study spans a broad range of domains demonstrating the ability and prospect of advanced deep learning (DL) and natural language processing (NLP) in mitigating the spread of fake news. Thus, high accuracy in fake news detection presented in the study may serve as a reliable guideline that would enable social media platforms, news outlets and even regulatory bodies to minimize fake news dissemination influence. The model's performance in examining complex features of linguistic phenomena and interpreting contextual information implies the possibility of a highly valuable role in Facebook's system of real-time AI-based content moderation. It could thus support endeavors in cultivating reliable media, protecting the public sphere, and avoiding a polarization of society. In addition, the study creates room for enhancing cybersecurity by minimizing the exploitation of fake news in phishing, social engineering, and other related vices. In academic and research contexts, the findings suggest considering the more complex models for better modelling of fake news that include multimodal data and new techniques used for spreading fake news. Furthermore, in practical aspects fundamental to public relations and marketing domains, where it is advisable to present accurate information to the targeted population, to enhance the loyalty and trust of consumers, it is also applicable. In sum, the findings of this study provide evidence in support of deploying state-of-art AI technologies for fighting misinformation to enhance the level of a digital community.

6.3 Recommendations for Future Research

Investigating numerous important areas can help future studies in false news detection enhance the performance, applicability, and flexibility of detection systems to the dynamic character of disinformation. These developments will enable the development of more dependable, efficient, and all-encompassing solutions for addressing the problem of false news across many platforms and circumstances.

- **Incorporate Multimodal Data:** Future work should try to combine text, images, videos, and metadata to make the detection models stronger. The integration of these modalities can improve detection quality and capture a broader context.
- **Real-Time Detection Systems:** Research should focus on lightweight and efficient models that can be deployed for real-time identification in social media and news feeds. These models should support large-scale data processing and timely responses.
- **Addressing Multilingual Challenges:** Future models must be able to detect false material in many languages and dialects, expanding their applicability globally.
- **Explainability and Transparency:** Future models should incorporate explainable AI practices to promote clarity in decision-making, essential for building trust in automation.
- **Ethical Considerations:** Issues such as dataset bias, fairness, and privacy must be addressed. Preventing ideological bias and ensuring inclusiveness in learning processes is crucial.

- **User Behavior and Social Dynamics:** Including user behavior and social network data can lead to more effective fake news detection. Understanding how misinformation spreads and how users interact with it can refine model strategies.
- **Collaboration with Social Media Platforms:** Research should explore integrating detection models in real-world environments in partnership with platforms to facilitate flagging, disputing, and limiting fake news spread.
- **Combatting Deepfakes and Synthetic Misinformation:** Combining NLP with computer vision approaches to detect manipulated images and videos is vital. Hybrid models could lead to more comprehensive multimedia fake content detection systems.
- **Practical deployment strategies:** For integrating Roberta into social media moderation pipelines. This includes optimizing for scalability and latency, potentially by combining Roberta with lightweight pre-filters or using it for batch processing. We also plan to address policy and regulatory considerations, such as ensuring transparency, enabling user appeals, and aligning with frameworks like the EU Digital Services Act. Furthermore, we recognize the importance of adversarial robustness—malicious actors may attempt to evade detection through subtle rewording or coded language. To mitigate this, we intend to investigate adversarial training and data augmentation methods to strengthen the model's resilience against such attacks.

6.4 Conclusion

In conclusion, the findings of this research support the use of the integrated Roberta sequence classifier model for fake news identification in multiple social media platforms. This study emphasizes the importance of different deep learning methodologies in particular Roberta in improving the efficiency and effectiveness of the fake news detection models. On the PolitiFact as well as Gossip Cop dataset, the proposed model had high accuracy of 93.55% and 94.19% respectively. Moreover, in the other evaluation parameters, the model performed well in precision where it achieved 94.12 % for PolitiFact and 97.55% for Gossip Cop, recall, where it got 94.12% for PolitiFact and 90.81% for Gossip Cop and F1-score where it got 94.12 % for PolitiFact and 94.06% for Gossip Cop, all of As expected, the proposed Roberta model yielded promising results in all aspects and surpassed the baseline models such as TChecker, dEFEND, and TCNN-URG that were used in our experiments for real-life implementation. Based on the research, deep learning models propounded improved solutions especially with the Roberta in the detection of fake news, which shows that the task of containing fake news is protrusible. Possible future work includes the improvement of the presented model for larger and more complex data sets and optimization of the detection process in order to provide real time detection across multiple platforms. Hence, this work can help academics to expand their knowledge and provide practitioners with effective ways of dealing with negative consequences of fake news.

REFERENCES

Abbas, Q., Zeshan, M.U. and Asif, M. (2022) ‘A CNN-RNN Based Fake News Detection Model Using Deep Learning’, in *Proceedings - 2022 International Seminar on Computer Science and Engineering Technology, SCSET 2022*. Available at: https://doi.org/10.1109/SCSET55041.2022.00019.

Abdulrahman, A. and Baykara, M. (2020) ‘Fake News Detection Using Machine Learning and Deep Learning Algorithms’, in *3rd International Conference on Advanced Science and Engineering, ICOASE 2020*. Available at: https://doi.org/10.1109/ICOASE51841.2020.9436605.

Abedalla, A., Al-Sadi, A. and Abdullah, M. (2019) ‘A closer look at fake news detection: A deep learning perspective’, in *ACM International Conference Proceeding Series*. Available at: https://doi.org/10.1145/3369114.3369149.

Abel Nyambe Mushiba, D.G.S. (2020) ‘FAKE NEWS DETECTION SYSTEM USING MACHINE LEARNING ALGORITHMS’, *International Journal of Creative Research Thoughts (IJCRT)*, 10(8), pp. 1–11.

Adedoyin, F. and Mariyappan, B. (2022) ‘Fake News Detection using Machine Learning Algorithms and Recurrent Neural Networks’, *Machine Learning Paradigm for Internet of Things Applications*, pp. 181–207. Available at: https://doi.org/10.31124/advance.20751379.v1.

Agarwal, A. *et al.* (2020) ‘Fake News Detection Using a Blend of Neural Networks: An Application of Deep Learning’, *SN Computer Science* [Preprint]. Available at: https://doi.org/10.1007/s42979-020-00165-4.

Ahmad, I. *et al.* (2020) ‘Fake News Detection Using Machine Learning Ensemble Methods’, *Complexity* [Preprint]. Available at: https://doi.org/10.1155/2020/8885861.

Ahmad, T. *et al.* (2022) ‘Efficient Fake News Detection Mechanism Using Enhanced Deep Learning Model’, *Applied Sciences (Switzerland)* [Preprint]. Available at: https://doi.org/10.3390/app12031743.

Ahmed, A.A.A., Aljabouh, A. and Praveen Kumar Donepudi, M.S.C. (2021) ‘Detecting Fake News using Machine Learning: A Systematic Literature Review’, *Psychology and Education Journal*, 58(1), pp. 1932–1939. Available at: https://doi.org/10.17762/pae.v58i1.1046.

Ahmed, S., Hinkelmann, K. and Corradini, F. (2022) ‘Development of Fake News Model using Machine Learning through Natural Language Processing’, 14(12), pp. 454–460.

Aïmeur, E., Amri, S. and Brassard, G. (2023) ‘Fake news, disinformation and misinformation in social media: a review’, *Social Network Analysis and Mining* [Preprint]. Available at: https://doi.org/10.1007/s13278-023-01028-5.

Akinduko, A.A., Mirkes, E.M. and Gorban, A.N. (2016) ‘SOM: Stochastic initialization versus principal components’, *Information Sciences*, 364–365, pp. 213–221. Available at: https://doi.org/10.1016/j.ins.2015.10.013.

Akram, B.A., Zhou, R. and Dong, F. (2024) ‘Fake News Classification Methodology With Enhanced BERT’, *IEEE Access*, 12(November), pp. 164491–164502. Available at: https://doi.org/10.1109/ACCESS.2024.3491376.

Al-alshaqi, M. and Liu, D.B.R. and C. (2024) ‘Ensemble Techniques for Robust Fake News Detection: Integrating Transformers, Natural Language Processing, and Machine Learning’, *sensors Article*, 24(18)(6062), pp. 1–13. Available at: https://doi.org/https://doi.org/10.3390/s24186062.

Al-tarawneh, M.A.B. *et al.* (2024) ‘Enhancing Fake News Detection with Word Embedding : A Machine Learning and Deep Learning Approach’, pp. 1–25.

Alarfaj, F.K. and Khan, J.A. (2023) ‘Deep Dive into Fake News Detection: Feature-Centric Classification with Ensemble and Deep Learning Methods’, *Algorithms*, 16(11). Available at: https://doi.org/10.3390/a16110507.

Alawadh, H.M. *et al.* (2023) ‘Attention-Enriched Mini-BERT Fake News Analyzer Using the Arabic Language’, *Future Internet* [Preprint]. Available at: https://doi.org/10.3390/fi15020044.

Aldwairi, M. and Alwahedi, A. (2018) ‘Detecting fake news in social media networks’, *Procedia Computer Science*, 141(January 2018), pp. 215–222. Available at: https://doi.org/10.1016/j.procs.2018.10.171.

Alghamdi, J., Lin, Y. and Luo, S. (2022) ‘A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection’, *Information (Switzerland)* [Preprint]. Available at: https://doi.org/10.3390/info13120576.

Ali, H. *et al.* (2021) ‘All Your Fake Detector Are Belong to Us: Evaluating Adversarial Robustness of Fake-News Detectors Under Black-Box Settings’, *IEEE Access*, 9, pp. 81678–81692. Available at: https://doi.org/10.1109/ACCESS.2021.3085875.

Ali, L. *et al.* (2021) ‘Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures’, *Sensors*, 21, p. 1688. Available at: https://doi.org/10.3390/s21051688.

Allcott, H. and Gentzkow, M. (2017) ‘Social media and fake news in the 2016 election’, *Journal of Economic Perspectives* [Preprint]. Available at: https://doi.org/10.1257/jep.31.2.211.

Allington, D. *et al.* (2021) ‘Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency’, *Psychological Medicine,* 51(10), pp. 1763–1769. Available at: https://doi.org/10.1017/S003329172000224X.

Altheneyan, A. and Alhadlaq, A. (2023) ‘Big Data ML-Based Fake News Detection Using Distributed Learning’, *IEEE Access*, 11(March), pp. 29447–29463. Available at: https://doi.org/10.1109/ACCESS.2023.3260763.

Alzubaidi, L. *et al.* (2021) ‘Review of deep learning: concepts, CNN architectures, challenges, applications, future directions’, *Journal of Big Data*, 8(1), p. 53. Available at: https://doi.org/10.1186/s40537-021-00444-8.

Amer, E., Kwak, K.S. and El-Sappagh, S. (2022) ‘Context-Based Fake News Detection Model Relying on Deep Learning Models’, *Electronics (Switzerland)*, 11(8), pp. 1–13. Available at: https://doi.org/10.3390/electronics11081255.

Angin, M. *et al.* (2022) ‘A RoBERTa Approach for Automated Processing of Sustainability Reports’, *Sustainability (Switzerland)* [Preprint]. Available at: https://doi.org/10.3390/su142316139.

Anjali Gangan, Sayali Wagh, Wajida Siddiqui, D.M.D. (2021) ‘Fake News Detection Using Machine Learning’, *International Journal of Research Publication and Reviews*, 2(4), pp. 562–564.

Al Asaad, B. and Erascu, M. (2018) ‘A tool for fake news detection’, *Proceedings - 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2018*, pp. 379–386. Available at: https://doi.org/10.1109/SYNASC.2018.00064.

Ashraf, N. *et al.* (2021) ‘CIC at CheckThat! 2021: Fake news detection using machine learning and data augmentation’, *CEUR Workshop Proceedings*, 2936, pp. 446–454.

Aslam, N. *et al.* (2021) ‘Fake Detect: A Deep Learning Ensemble Model for Fake News Detection’, *Complexity* [Preprint]. Available at: https://doi.org/10.1155/2021/5557784.

Ayenew, A. *et al.* (2024) ‘An Evolutionary Amharic Fake News Detection Based on ML and Deep Learning Approach’, *Journal of Electrical Systems*, 20(3), pp. 1–11.

Baashirah, R. (2024) ‘Zero-Shot Automated Detection of Fake News: An Innovative Approach (ZS-FND)’, *IEEE Access*, PP, p. 1. Available at: https://doi.org/10.1109/ACCESS.2024.3462151.

Balshetwar, S. V., Abilash, R.S. and Dani Jermisha, R. (2023) ‘Fake news detection in social media based on sentiment analysis using classifier techniques’, *Multimedia Tools and Applications* [Preprint]. Available at: https://doi.org/10.1007/s11042-023-14883-3.

Bangyal, W.H. *et al.* (2021) ‘Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches’, 2021. Available at: https://doi.org/10.1155/2021/5514220.

Batchelor, O. (2017) ‘Getting Out the Truth: The Role of Libraries in the Fight against Fake News’, *Reference Services Review*, 45. Available at: https://doi.org/10.1108/RSR-03-2017-0006.

Bauskar, S. *et al.* (2019) ‘Natural Language Processing based Hybrid Model for Detecting Fake News Using Content-Based Features and Social Features’, *International Journal of Information Engineering and Electronic Business*, 11(4), pp. 1–10. Available at: https://doi.org/10.5815/ijieeb.2019.04.01.

Baym, G. (2005) ‘The Daily Show: Discursive Integration and the Reinvention of Political Journalism’, *Political Communication*, 22(3), pp. 259–276. Available at: https://doi.org/10.1080/10584600591006492.

Belorkar, A. *et al.* (2020) ‘Interactive Data Visualization with Python Plotly’, *2020, Second edition.*, p. 343.

Bengio, Y. (2013) ‘Deep Learning of Representations: Looking Forward’, in A.-H. Dediu et al. (eds) *Statistical Language and Speech Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–37.

Bermes, A. (2021) 'Information overload and fake news sharing: A transactional stress perspective exploring the mitigating role of consumers' resilience during COVID-19.', *Journal of Retailing and Consumer Services*, p. 102555. Available at: https://doi.org/10.1016/j.jretconser.2021.102555.

Bhat, V. (2023) 'An Overview Of Essential Python Libraries And Tools For Data Science- A Review Paper', *International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org*, 5(2), pp. 85–90. Available at: https://doi.org/10.33093/jetap.2023.5.2.10.

Bishop, C.M. and Lasserre, J. (2007) 'Bayesian Statistics: Generative or Discriminative ? Getting the Best of Both Worlds', *Statistics* [Preprint].

Bondielli, A. and Marcelloni, F. (2019) 'A survey on fake news and rumour detection techniques', *Information Sciences*, 497, pp. 38–55. Available at: https://doi.org/https://doi.org/10.1016/j.ins.2019.05.035.

Brewer, P.R., Young, D.G. and Morreale, M. (2013) 'The Impact of Real News about "Fake News": Intertextual Processes and Political Satire', *International Journal of Public Opinion Research*, 25(3), pp. 323–343. Available at: https://doi.org/10.1093/ijpor/edt015.

Broersma, M. and Graham, T. (2013) 'Twitter as a News Source – How Dutch And British Newspapers Used Tweets in Their News Coverage', *Journalism Practice*, 7, pp. 446–464. Available at: https://doi.org/10.1080/17512786.2013.802481.

Buhmann, M.D. (2000) 'Radial basis functions', *Acta Numerica*, 9, pp. 1–38. Available at: https://doi.org/10.1017/S0962492900000015.

Caetano, J.A. *et al.* (2018) 'Analyzing and characterizing political discussions in

WhatsApp public groups’, *CoRR*, abs/1804.0.

Canini, K., Suh, B. and Pirolli, P. (2011) ‘Finding Credible Information Sources in Social Networks Based on Content and Social Structure’, in, pp. 1–8. Available at: https://doi.org/10.1109/PASSAT/SocialCom.2011.91.

Chandel, M. *et al.* (2022) ‘A Study on Machine Learning and Python’s Framework’, *International Journal of Computer Sciences and Engineering Open Access Survey Paper*, 10(5).

Chang, L.-J. (2024) ‘Comparison of Machine Learning and Deep Learning Algorithms in Detecting Fake News’, *Proceedings of the 28th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2024*, (Wmsci), pp. 203–209. Available at: https://doi.org/10.54808/wmsci2024.01.203.

Chen, K. (2015) ‘Deep and Modular Neural Networks’, in J. Kacprzyk and W. Pedrycz (eds) *Springer Handbook of Computational Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 473–494. Available at: https://doi.org/10.1007/978-3-662-43505-2_28.

Chen, Y., Conroy, N.J. and Rubin, V.L. (2015) ‘Misleading Online Content: Recognizing Clickbait as “False News”’, in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*. New York, NY, USA: Association for Computing Machinery (WMDD ’15), pp. 15–19. Available at: https://doi.org/10.1145/2823465.2823467.

Collen, D.R., Nyandoro, L.K. and Zvarevashe, K. (2022) ‘Fake News Detection using 5L-CNN’, in *2022 1st Zimbabwe Conference of Information and Communication Technologies, ZCICT 2022*. Available at: https://doi.org/10.1109/ZCICT55726.2022.10045981.

Dhanuka, O. and Tiwari, D.S. (2024) ‘Fake News Detection Using Machine Learning’, *International Journal of Research Publication and Reviews*, 5(1), pp. 2657–2664. Available at: https://doi.org/10.55248/gengpi.5.0124.0264.

Dhar, M. (2022) 'Fake News Detection Using Machine Learning And Deep Learning', in, pp. 1–6.

Dr.Jayakumar, Niket kumar, N.H. (2021) 'Fake News Detection using ML', *Turkish Journal of Computer and Mathematics Education*, 12(11), pp. 1–5. Available at: https://doi.org/10.17762/turcomat.v12i11.6331.

E Almandouh, M. *et al.* (2024) 'Ensemble based high performance deep learning models for fake news detection', *Scientific reports*, 14(1), p. 26591. Available at: https://doi.org/10.1038/s41598-024-76286-0.

Fan, C. *et al.* (2021) 'A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data', *Frontiers in Energy Research*, 9. Available at: https://doi.org/10.3389/fenrg.2021.652801.

Farooq, M.S. *et al.* (2023) 'Fake news detection in Urdu language using machine learning', *PeerJ Computer Science*, 9, pp. 1–20. Available at: https://doi.org/10.7717/peerj-cs.1353.

Fernandez-Delgado, M. *et al.* (2014) 'Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?', *Journal of Machine Learning Research*, 15, pp. 3133–3181.

Fouad, K.M., Sabbeh, S.F. and Medhat, W. (2022) 'Arabic fake news detection using deep learning', *Computers, Materials and Continua*, 71(2), pp. 3647–3665. Available at: https://doi.org/10.32604/cmc.2022.021449.

Fu, Z. and Jiang, J. (2019) 'Experimental Study on Machine Learning with Approximation to Data Streams', *2019 6th International Conference on Internet of Things: Systems, Management and Security, IOTSMS 2019*, pp. 561–566.

Available at: https://doi.org/10.1109/IOTSMS48152.2019.8939225.

GabAllah, N., Sharara, H. and Rafea, A. (2023) 'TChecker: A Content Enrichment Approach for Fake News Detection on Social Media', *Applied Sciences (Switzerland)*, 13(24). Available at: https://doi.org/10.3390/app132413070.

Garg, A., Saudagar, S. and Gupta, A.S.A.L.G.G. (2024) 'Identifying Fake News On ISOT Data Using Stemming Method With A Subdomain Of AI Algorithms', 21(S6), pp. 775–787.

Garg, S. and Sharma, D.K. (2022) 'Fake News Classification via CNN', in *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*. Available at: https://doi.org/10.1109/SMART55829.2022.10047560.

Garimella, K. and Tyson, G. (2018) 'WhatApp Doc? A First Look at WhatsApp Public Group Data', *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1). Available at: https://doi.org/10.1609/icwsm.v12i1.14989.

Girgis, S., Amer, E. and Gadallah, M. (2018) 'Deep Learning Algorithms for Detecting Fake News in Online Text', in *Proceedings - 2018 13th International Conference on Computer Engineering and Systems, ICCES 2018*. Available at: https://doi.org/10.1109/ICCES.2018.8639198.

Goldani, M.H., Momtazi, S. and Safabakhsh, R. (2021) 'Detecting fake news with capsule neural networks', *Applied Soft Computing* [Preprint]. Available at: https://doi.org/10.1016/j.asoc.2020.106991.

Gravanis, G. *et al.* (2019) 'Behind the cues: A benchmarking study for fake news detection', *Expert Systems with Applications* [Preprint]. Available at: https://doi.org/10.1016/j.eswa.2019.03.036.

Guille, A. *et al.* (2013) 'Information diffusion in online social networks: a survey', *SIGMOD Rec.*, 42(2), pp. 17–28. Available at: https://doi.org/10.1145/2503792.2503797.

GÜLER, G. and GÜNDÜZ, S. (2023) ‘Deep Learning Based Fake News Detection on Social Media’, *International Journal of Information Security Science*, 12(2), pp. 1–21. Available at: https://doi.org/10.55859/ijiss.1231423.

Gundapu, S. and Mamidi, R. (2021) ‘Transformer based Automatic COVID-19 Fake News Detection System’, pp. 1–12. Available at: https://doi.org/https://doi.org/10.48550/arXiv.2101.00180.

Guo, Y., Lamaazi, H. and Mizouni, R. (2022) ‘Smart Edge-based Fake News Detection using Pre-trained BERT Model’, in *International Conference on Wireless and Mobile Computing, Networking and Communications*. Available at: https://doi.org/10.1109/WiMob55322.2022.9941689.

Gupta, A., Lamba, H. and Kumaraguru, P. (2013) ‘Analyzing fake content on Twitter’, in *2013 APWG eCrime Researchers Summit*, pp. 1–12. Available at: https://doi.org/10.1109/eCRS.2013.6805772.

Gupta, S. and Meel, P. (2021) *Inventive Communication and Computational Technologies*, *Lecture Notes in Networks and Systems*. Available at: https://doi.org/https://doi.org/10.1007/978-981-15-7345-3_13.

H.Shanthi *et al.* (2022) ‘FAKE NEWS DETECTION USING MACHINE LEARNING’, 54(2), pp. 1011–1020.

Hamed, S.K., Ab Aziz, M.J. and Yaakub, M.R. (2023) ‘Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users’ Comments’, *Sensors (Basel, Switzerland)*, 23(4). Available at: https://doi.org/10.3390/s23041748.

Hamid, A. *et al.* (2020) ‘Fake news detection in social media using graph neural networks and NLP techniques: A COVID-19 use-case', *CEUR Workshop Proceedings*, 2882, pp. 19–21.

Hangloo, S. and Arora, B. (2022) 'Content-based fake news detection using deep learning techniques: Analysis, challenges and possible solutions', in *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, pp. 411–417. Available at: https://doi.org/10.1109/CCiCT56684.2022.00079.

Hansrajh, A., Adeliyi, T.T. and Wing, J. (2021) 'Detection of Online Fake News Using Blending Ensemble Learning', *Scientific Programming* [Preprint]. Available at: https://doi.org/10.1155/2021/3434458.

Hashmi, E. *et al.* (2024) 'Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI', *IEEE Access*, 12(March), pp. 44462–44480. Available at: https://doi.org/10.1109/ACCESS.2024.3381038.

Hatwar, N.. *et al.* (2022) 'Fake News Detection Using Machine Learning', *Lecture Notes in Networks and Systems*, 4(6), pp. 2579–2583. Available at: https://doi.org/10.1007/978-981-99-0769-4_41.

He, K. *et al.* (2016) 'Deep residual learning for image recognition', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Available at: https://doi.org/10.1109/CVPR.2016.90.

Hegde, A. and Shashirekha, H.L. (2021) 'Urdu Fake News Detection Using Ensemble of Machine Learning Models', *CEUR Workshop Proceedings*, 3159, pp. 1190–1198.

Hisham, M., Hasan, R. and Hussain, S. (2023) 'An Innovative Approach for Fake News Detection using Machine Learning', *Sir Syed University Research Journal of Engineering & Technology*, 13(1), pp. 115–124. Available at: https://doi.org/10.33317/ssurj.565.

Hochreiter, S. and Schmidhuber, J. (1997) ‘Long Short-Term Memory’, *Neural Computation* [Preprint]. Available at: https://doi.org/10.1162/neco.1997.9.8.1735.

Hossain, E. *et al.* (2021) ‘A Study towards Bangla Fake News Detection Using Machine Learning and Deep Learning’, *Springer, Singapore*, 1408, pp. 1–15. Available at: https://doi.org/https://doi.org/10.1007/978-981-16-5157-1_7.

Hosseini, M. *et al.* (2023) ‘Interpretable fake news detection with topic and deep variational models’, *Online Social Networks and Media*, 36, pp. 1–12. Available at: https://doi.org/10.1016/j.osnem.2023.100249.

Isa, S.M., Nico, G. and Permana, M. (2022) ‘Indobert for Indonesian Fake News Detection’, *ICIC Express Letters*, 16(3), pp. 289–297. Available at: https://doi.org/10.24507/icicel.16.03.289.

Islam, N. *et al.* (2021) ‘Ternion: An autonomous model for fake news detection’, *Applied Sciences (Switzerland)* [Preprint]. Available at: https://doi.org/10.3390/app11199292.

Jadhav, P.A. *et al.* (2022) ‘Enhancing Fake News Detection System Using Python & Machine Learning’, *International Journal of Creative Research Thoughts (IJCRT)*, 10(1), pp. 915–918.

Jadhav, S.S. and Thepade, S.D. (2019) ‘Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier’, *Applied Artificial Intelligence*, 33(12), pp. 1058–1068. Available at: https://doi.org/10.1080/08839514.2019.1661579.

Jawad, Z.A. and Obaid, A.J. (2023) ‘the Combination of Convolution Neural Networks and Deep Neural Networks for Fake News Detection’, *Journal of Engineering Science and Technology*, 18(1), pp. 814–826. Available at: https://doi.org/https://doi.org/10.48550/arXiv.2210.08331.

Jayakody, N., Mohammad, A. and Halgamuge, M.N. (2022) ‘Fake News Detection using a Decentralized Deep Learning Model and Federated Learning’, in *IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1–6. Available at: https://doi.org/10.1109/IECON49645.2022.9968358.

Jaybhaye, S.M. *et al.* (2023) ‘Fake News Detection using LSTM based deep learning approach’, *ITM Web of Conferences* [Preprint]. Available at: https://doi.org/10.1051/itmconf/20235603005.

Jiang, T. *et al.* (2020) ‘Fake News Detection using Deep Recurrent Neural Networks’, in *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020*. Available at: https://doi.org/10.1109/ICCWAMTIP51612.2020.9317325.

Jiang, T. *et al.* (2021) ‘A Novel Stacking Approach for Accurate Detection of Fake News’, *IEEE Access*, 9, pp. 22626–22639. Available at: https://doi.org/10.1109/ACCESS.2021.3056079.

Joseph, V.R. (2022) ‘Optimal ratio for data splitting’, *Statistical Analysis and Data Mining* [Preprint]. Available at: https://doi.org/10.1002/sam.11583.

Jouhar, J. *et al.* (2024) ‘Fake News Detection using Python and Machine Learning’, *Procedia Computer Science*, 233(2023), pp. 763–771. Available at: https://doi.org/10.1016/j.procs.2024.03.265.

Kaliyar, R.K. *et al.* (2020) ‘FNDNet – A deep convolutional neural network for fake news detection’, *Cognitive Systems Research* [Preprint]. Available at: https://doi.org/10.1016/j.cogsys.2019.12.005.

Kaliyar, R.K., Goswami, A. and Narang, P. (2021) ‘FakeBERT: Fake news detection in social media with a BERT-based deep learning approach’, *Multimedia Tools and Applications* [Preprint]. Available at: https://doi.org/10.1007/s11042-020-10183-

2.

Kanagavalli, N. and Priya, S.B. (2022) ‘Social Networks Fake Account and Fake News Identi fi cation with Reliable Deep Learning’. Available at: https://doi.org/10.32604/iasc.2022.022720.

Kaur, S., Kumar, P. and Kumaraguru, P. (2020) ‘Automating fake news detection system using multi-level voting model’, *Soft Computing*, 24(12), pp. 9049–9069. Available at: https://doi.org/10.1007/s00500-019-04436-y.

Kausar, N., Alikhan, A. and Sattar, M. (2022) ‘Towards better representation learning using hybrid deep learning model for fake news detection’, 12(165), pp. 1–11. Available at: https://doi.org/https://doi.org/10.1007/s13278-022-00986-6.

Khadka, N. (2019) ‘General Machine Learning Practices Using Python’.

Khalil, A. *et al.* (2021) ‘Detecting Arabic Fake News Using Machine Learning’, in *2021 2nd International Conference on Intelligent Data Science Technologies and Applications, IDSTA 2021*. Available at: https://doi.org/10.1109/IDSTA53674.2021.9660811.

Khan, M. *et al.* (2020) ‘Fake News Identification on Social Media’, *International Journal of Engineering Research and*, V9. Available at: https://doi.org/10.17577/IJERTV9IS010183.

Khan, R.H. *et al.* (2022) ‘Improved Fake News Detection Method based on Deep Learning and Comparative Analysis with other Machine Learning approaches’, in *8th International Conference on Engineering and Emerging Technologies, ICEET 2022*. Available at: https://doi.org/10.1109/ICEET56468.2022.10007214.

Khan, S. *et al.* (2022) ‘Detecting COVID-19-Related Fake News Using Feature Extraction’, *Frontiers in Public Health*, 9(January), pp. 1–9. Available at: https://doi.org/10.3389/fpubh.2021.788074.

Khanam, Z. *et al.* (2021) 'Fake News Detection Using Machine Learning Approaches', *IOP Conference Series: Materials Science and Engineering*, 1099(1), p. 012040. Available at: https://doi.org/10.1088/1757-899X/1099/1/012040.

Kitanovski, A., Toshevska, M. and Mirceva, G. (2023) 'DistilBERT and RoBERTa Models for Identification of Fake News', *2023 46th ICT and Electronics Convention, MIPRO 2023 - Proceedings*, (November), pp. 1102–1106. Available at: https://doi.org/10.23919/MIPRO57284.2023.10159740.

Kolev, V., Weiss, G. and Spanakis, G. (2022) 'FOREAL: RoBERTa Model for Fake News Detection based on Emotions', *International Conference on Agents and Artificial Intelligence*, 2(Icaart), pp. 429–440. Available at: https://doi.org/10.5220/0010873900003116.

Kousika, N. *et al.* (2021) 'A system for fake news detection by using supervised learning model for social media contents', in *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*. Available at: https://doi.org/10.1109/ICICCS51141.2021.9432096.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) 'ImageNet classification with deep convolutional neural networks', in *Advances in Neural Information Processing Systems*.

Kumar, A., Singh, J.P. and Singh, A.K. (2022) 'COVID-19 Fake News Detection Using Ensemble-Based Deep Learning Model', *IT Professional*, 24(2), pp. 32–37. Available at: https://doi.org/10.1109/MITP.2022.3149140.

Kumar, S. *et al.* (2020) 'Fake news detection using deep learning models: A novel approach', *Transactions on Emerging Telecommunications Technologies* [Preprint]. Available at: https://doi.org/10.1002/ett.3767.

Kushwaha, N.S. and Singh, P. (2022) 'Fake News Detection using Machine Learning: A

Comprehensive Analysis', *Journal of Management and Service Science (JMSS)*, 2(1), pp. 1–15. Available at: https://doi.org/10.54060/JMSS/002.01.001.

Lahby, M. and Yassine, M. (2022) *Combating Fake News with Computational Intelligence Techniques, Studies in Computational Intelligence*. Available at: https://doi.org/https://doi.org/10.1007/978-3-030-90087-8_7.

Lakshmi, J.V.N. (2018) 'Machine learning techniques using python for data analysis in performance evaluation', *International Journal of Intelligent Systems Technologies and Applications* [Preprint]. Available at: https://doi.org/10.1504/ijista.2018.10012853.

Lecun, Y. *et al.* (1998) 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, 86(11), pp. 2278–2324. Available at: https://doi.org/10.1109/5.726791.

Lecun, Y. and Bengio, Y. (1995) 'Convolutional Networks for Images, Speech, and Time-Series', in *The Handbook of Brain Theory and Neural Networks*.

LeCun, Y., Kavukcuoglu, K. and Farabet, C. (2010) 'Convolutional networks and applications in vision', in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. IEEE, pp. 253–256. Available at: https://doi.org/10.1109/ISCAS.2010.5537907.

Liu, Y. and Wu, Y.-F.B. (2020) 'FNED: A Deep Network for Fake News Early Detection on Social Media', *ACM Transactions on Information Systems*, 38(3), pp. 1–33. Available at: https://doi.org/10.1145/3386253.

Madani, M., Motameni, H. and Roshani, R. (2024) 'Fake News Detection Using Feature Extraction, Natural Language Processing, Curriculum Learning, and Deep Learning', *International Journal of Information Technology and Decision Making*, 23(3), pp. 1063–1098. Available at:

https://doi.org/10.1142/S0219622023500347.

Madhunitha, R. (2024) ‘Fake News Detection Using Machine Learning’, *International Journal for Research in Applied Science and Engineering Technology*, 12(7), pp. 345–351. Available at: https://doi.org/10.22214/ijraset.2024.63570.

Mahabub, A. (2020) ‘A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers’, *SN Applied Sciences*, 2(4), p. 525. Available at: https://doi.org/10.1007/s42452-020-2326-y.

Maham, S. *et al.* (2024) ‘ANN: adversarial news net for robust fake news classification’, *Scientific Reports*, 14(1), pp. 1–20. Available at: https://doi.org/10.1038/s41598-024-56567-4.

Mahara, G.S. and Gangele, S. (2022) ‘Fake news detection: A RNN-LSTM, Bi-LSTM based deep learning approach’, in *2022 IEEE 1st International Conference on Data, Decision and Systems, ICDDS 2022*. Available at: https://doi.org/10.1109/ICDDS56399.2022.10037403.

Mahmud, F.B. *et al.* (2022) ‘A comparative analysis of Graph Neural Networks and commonly used machine learning algorithms on fake news detection’, *Proceedings - 2022 7th International Conference on Data Science and Machine Learning Applications, CDMA 2022*, pp. 97–102. Available at: https://doi.org/10.1109/CDMA54072.2022.00021.

Mahmud, T. *et al.* (2024) ‘Integration of NLP and Deep Learning for Automated Fake News Detection’, *Proceedings - 2024 2nd International Conference on Inventive Computing and Informatics, ICICI 2024*, (September), pp. 398–404. Available at: https://doi.org/10.1109/ICICI62254.2024.00072.

Mallick, C., Mishra, S. and Senapati, M.R. (2023) ‘A cooperative deep learning model for fake news detection in online social networks’, *Journal of Ambient*

Intelligence and Humanized Computing [Preprint]. Available at: https://doi.org/10.1007/s12652-023-04562-4.

Martínez-Gallego, K., Álvarez-Ortiz, A.M. and Arias-Londoño, J.D. (2021) 'Fake News Detection in Spanish Using Deep Learning Techniques', pp. 1–10.

Matheven, A. and Kumar, B.V.D. (2022) 'Fake News Detection Using Deep Learning and Natural Language Processing', in *2022 9th International Conference on Soft Computing and Machine Intelligence, ISCMI 2022*. Available at: https://doi.org/10.1109/ISCMI56532.2022.10068440.

Meesad, P. (2021) 'Thai Fake News Detection Based on Information Retrieval , Natural Language Processing and Machine Learning', *SN Computer Science*, 2(6), pp. 1–17. Available at: https://doi.org/10.1007/s42979-021-00775-6.

Meese, J., Frith, J. and Wilken, R. (2020) 'COVID-19, 5G conspiracies and infrastructural futures.', *Media International Australia*, pp. 30–46. Available at: https://doi.org/10.1177/1329878X20952165.

Mir, Y. and sandhu, N. (2019) 'A Review on Deep Learning Architectures and Applications: Image Recognition'.

Mishra, A. and Sadia, H. (2023) 'A Comprehensive Analysis of Fake News Detection Models: A Systematic Literature Review and Current Challenges †', *Engineering Proceedings* [Preprint]. Available at: https://doi.org/10.3390/engproc2023059028.

Mohawesh, R. *et al.* (2023) 'Semantic graph based topic modelling framework for multilingual fake news detection', *AI Open* [Preprint]. Available at: https://doi.org/10.1016/j.aiopen.2023.08.004.

Mohsen, F. *et al.* (2024) 'Automated Detection of Misinformation : A Hybrid Approach for Fake News Detection', pp. 1–24.

Moisi, E.V. *et al.* (2024) 'Romanian Fake News Detection using Machine Learning and

Transformer- Based Approaches', pp. 0–21. Available at: https://doi.org/10.20944/preprints202411.1843.v1.

Monti, F. *et al.* (2019) 'Fake News Detection on Social Media using Geometric Deep Learning', pp. 1–15.

Mouratidis, D., Nikiforos, M.N. and Kermanidis, K.L. (2021) 'Deep learning for fake news detection in a pairwise textual input schema', *Computation*, 9(2), pp. 1–15. Available at: https://doi.org/10.3390/computation9020020.

Mujahid, M. *et al.* (2024) 'Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering', *Journal of Big Data*, 11(1), p. 87. Available at: https://doi.org/10.1186/s40537-024-00943-4.

Nadeem, M.I. *et al.* (2023) 'HyproBert: A Fake News Detection Model Based on Deep Hypercontext', *Symmetry*, 15(2), pp. 1–21. Available at: https://doi.org/10.3390/sym15020296.

Nagi, K. (2018) 'New Social Media and Impact of Fake News on Society by Kuldeep Nagi', *ICSSM Proceedings, July*, pp. 77–96.

Nath, K. *et al.* (2021) 'Study of Fake News Detection using Machine Learning and Deep Learning Classification Methods', in *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, pp. 434–438. Available at: https://doi.org/10.1109/RTEICT52294.2021.9573583.

Navlani, A. (2019) 'Text ANALYTICS for Beginners using NLTK', *Datacamp*, pp. 1–14.

Newman, N., Dutton, W. and Blank, G. (2014) 'Social Media and the News', in, pp. 135–148. Available at: https://doi.org/10.1093/acprof:oso/9780199661992.003.0009.

News, F. *et al.* (2022) 'Mc-DNN: Fake News Detection Using MultiChannel Deep Neural Networks', 18(1), pp. 1–20. Available at:

https://doi.org/10.4018/IJSWIS.295553.

Ng, A.Y. and Jordan, M.I. (2002) 'On discriminative vs. Generative classifiers: A comparison of logistic regression and naive bayes', in *Advances in Neural Information Processing Systems*.

Nickerson, R. (1998) 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises', *Review of General Psychology*, 2, pp. 175–220. Available at: https://doi.org/10.1037/1089-2680.2.2.175.

Nidha, A. *et al.* (2024) 'Fake News Detection Using Machine Learning and Web scraping', *International Research Journal of Engineering and Technology (IRJET)*, 11(5), pp. 1136–1143.

Nirav Shah, M. and Ganatra, A. (2022) 'A systematic literature review and existing challenges toward fake news detection models', *Social Network Analysis and Mining*, 12(1), p. 168. Available at: https://doi.org/10.1007/s13278-022-00995-5.

Nyow, N.X. and Chua, H.N. (2019) 'Detecting Fake News with Tweets' Properties', in *2019 IEEE Conference on Application, Information and Network Security (AINS)*, pp. 24–29. Available at: https://doi.org/10.1109/AINS47559.2019.8968706.

Olan, F. *et al.* (2024) 'Fake news on Social Media: the Impact on Society', *Information Systems Frontiers* [Preprint]. Available at: https://doi.org/10.1007/s10796-022-10242-z.

Ouassil, M.A. *et al.* (2022) 'A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model', *International Journal of Advanced Computer Science and Applications*, 13(10), pp. 525–534. Available at: https://doi.org/10.14569/IJACSA.2022.0131061.

Paas, F. *et al.* (2003) 'Cognitive load measurement as a means to advance cognitive load theory', *Educational Psychologist* [Preprint]. Available at:

https://doi.org/10.1207/S15326985EP3801_8.

Palani, B., Elango, S. and Viswanathan K, V. (2022) 'CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT', *Multimedia Tools and Applications*, 81(4), pp. 5587–5620. Available at: https://doi.org/10.1007/s11042-021-11782-3.

Pandey, S. *et al.* (2022) 'Fake News Detection from Online media using Machine learning Classifiers', *Journal of Physics: Conference Series*, 2161(1), p. 12027. Available at: https://doi.org/10.1088/1742-6596/2161/1/012027.

Patkar, U. *et al.* (2022) 'PYTHON FOR WEB DEVELOPMENT', *International Journal of Computer Science and Mobile Computing* [Preprint]. Available at: https://doi.org/10.47760/ijcsmc.2022.v11i04.006.

Pérez-Rosas, V. *et al.* (2018) 'Automatic detection of fake news', *COLING 2018 - 27th International Conference on Computational Linguistics, Proceedings*, pp. 3391–3401.

Prabhakar, E. (2024) 'A Novel Voting Ensemble Stacking Approach for Fake News Detection', 13(4).

Prachi, N.N. *et al.* (2022) 'Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms', *Journal of Advances in Information Technology*, 13(6), pp. 652–661. Available at: https://doi.org/10.12720/jait.13.6.652-661.

Pranay Patil, Abrar Khan, M.M. (2021) 'Fake News Detection Using ml', *International Research Journal of Modernization in Engineering Technology and Science*, 3(5), pp. 562–563.

Prathanrat, P. and Polprasert, C. (2018) 'Performance Prediction of Jupyter Notebook in JupyterHub using Machine Learning', in *2018 International Conference on*

Intelligent Informatics and Biomedical Sciences, ICIIBMS 2018. Available at: https://doi.org/10.1109/ICIIBMS.2018.8550030.

Pre-proofs, J. (2020) 'Linguistic Feature Based Learning Model for Fake News', *Expert Systems With Applications*, p. 114171. Available at: https://doi.org/10.1016/j.eswa.2020.114171.

Priya, A. and Kumar, A. (2021) 'Deep Ensemble Approach for COVID-19 Fake News Detection from Social Media', in *Proceedings of the 8th International Conference on Signal Processing and Integrated Networks, SPIN 2021*. Available at: https://doi.org/10.1109/SPIN52536.2021.9565958.

Qawasmeh, E., Tawalbeh, M. and Abdullah, M. (2019) 'Automatic Identification of Fake News Using Deep Learning', *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, (December), pp. 383–388. Available at: https://doi.org/10.1109/SNAMS.2019.8931873.

Qian, F. *et al.* (2018) 'Neural user response generator: Fake news detection with collective user intelligence', *IJCAI International Joint Conference on Artificial Intelligence*, 2018-July, pp. 3834–3840. Available at: https://doi.org/10.24963/ijcai.2018/533.

Radford, A., Metz, L. and Chintala, S. (2016) 'Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks'.

Rafique, A. *et al.* (2022) 'Comparative analysis of machine learning methods to detect fake news in an Urdu language corpus', *PeerJ Computer Science*, 8, pp. 1–20. Available at: https://doi.org/10.7717/peerj-cs.1004.

Rahman, M.S., Ashraf, F. Bin and Kabir, M.R. (2023) 'An Efficient Deep Learning Technique for Bangla Fake News Detection', in. Available at: https://doi.org/10.1109/iccit57492.2022.10055636.

Rahmawati, A., Alamsyah, A. and Romadhony, A. (2022) ‘Hoax News Detection Analysis using IndoBERT Deep Learning Methodology’, *2022 10th International Conference on Information and Communication Technology, ICoICT 2022*, (April), pp. 368–373. Available at: https://doi.org/10.1109/ICoICT55009.2022.9914902.

Rajathi M (2021) ‘A Study on machine learning algorithms on big data analytics’.

Rampurkar, V. and Thirupurasundari, D.R. (2024) ‘An Approach towards Fake News Detection using Machine Learning Techniques’, *INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING*, 12(3), pp. 1–7.

Rao, S., Verma, A.K. and Bhatia, T. (2021) ‘A review on social spam detection: Challenges, open issues, and future directions’, *Expert Systems with Applications*, 186, p. 115742. Available at: https://doi.org/https://doi.org/10.1016/j.eswa.2021.115742.

Rasel, R.I. *et al.* (2022) ‘Bangla Fake News Detection using Machine Learning, Deep Learning and Transformer Models’, *Proceedings of 2022 25th International Conference on Computer and Information Technology, ICCIT 2022*, (July 2023), pp. 959–964. Available at: https://doi.org/10.1109/ICCIT57492.2022.10055592.

Reddy, H. *et al.* (2020) ‘Text-mining-based Fake News Detection Using Ensemble Methods’, *International Journal of Automation and Computing*, 17(2), pp. 210–221. Available at: https://doi.org/10.1007/s11633-019-1216-5.

Reis, J.C.S. *et al.* (2019) ‘Explainable machine learning for fake news detection’, *WebSci 2019 - Proceedings of the 11th ACM Conference on Web Science*, pp. 17–26. Available at: https://doi.org/10.1145/3292522.3326027.

Ren, J.S.J. and Xu, L. (2015) ‘On Vectorization of Deep Convolutional Neural Networks for Vision Tasks’, *CoRR*, abs/1501.0.

Roh, Y., Heo, G. and Whang, S.E. (2021) ‘A Survey on Data Collection for Machine Learning: A Big Data-AI Integration Perspective’, *IEEE Transactions on Knowledge and Data Engineering*, 33(4), pp. 1328–1347. Available at: https://doi.org/10.1109/TKDE.2019.2946162.

Ross, L. and Ward, A. (1996) ‘Naive Realism: Implications for Social Conflict and Misunderstanding’, in, pp. 103–135.

Roy, A. *et al.* (2018) ‘A Deep Ensemble Framework for Fake News Detection and Classification’.

Ruchansky, N., Seo, S. and Liu, Y. (2017) ‘CSI: A hybrid deep model for fake news detection’, *International Conference on Information and Knowledge Management, Proceedings*, Part F1318, pp. 797–806. Available at: https://doi.org/10.1145/3132847.3132877.

Sadeghi, F., Bidgoly, A.J. and Amirkhani, H. (2020) ‘Fake news detection on social media using a natural language inference approach’, *Multimedia Tools and Applications*, 81(23), pp. 33801–33821. Available at: https://doi.org/10.1007/s11042-022-12428-8.

Saeed, A. and Al Solami, E. (2023) ‘Fake News Detection Using Machine Learning and Deep Learning Methods’, *Computers, Materials and Continua*, 77, pp. 2079–2096. Available at: https://doi.org/10.32604/cmc.2023.030551.

Saikh, T. *et al.* (2020) ‘A Deep Learning Approach for Automatic Detection of Fake News’.

Saleh, H., Alharbi, A. and Alsamhi, S.H. (2021) ‘OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection’, *IEEE Access* [Preprint]. Available at: https://doi.org/10.1109/ACCESS.2021.3112806.

Salh, D.A. and Nabi, R.M. (2023) ‘Kurdish Fake News Detection Based on Machine

Learning Approaches', *Passer Journal of Basic and Applied Sciences*, 5(2), pp. 262–271. Available at: https://doi.org/10.24271/PSR.2023.380132.1226.

Sarnovský, M., Maslej-kre, V. and Ivancová, K. (2022) 'Fake News Detection Related to the COVID-19 in Slovak Language Using Deep Learning Methods', 19(2), pp. 43–57.

Schmidhuber, J. (2015) 'Deep Learning in neural networks: An overview', *Neural Networks* [Preprint]. Available at: https://doi.org/10.1016/j.neunet.2014.09.003.

Segura-Bedmar, I. and Alonso-Bartolome, S. (2022) 'Multimodal Fake News Detection', *Information (Switzerland)*, 13(6). Available at: https://doi.org/10.3390/info13060284.

Semanjski, I.C. (2023) 'Chapter 5 - Data analytics', in I.C. Semanjski (ed.) *Smart Urban Mobility*. Elsevier, pp. 121–170. Available at: https://doi.org/https://doi.org/10.1016/B978-0-12-820717-8.00008-7.

Senhadji, S. and Ahmed, R.A.S. (2022) 'Fake news detection using naïve Bayes and long short term memory algorithms', *IAES International Journal of Artificial Intelligence*, 11(2), pp. 746–752. Available at: https://doi.org/10.11591/ijai.v11.i2.pp746-752.

Series, C. (2021) 'Fake News Detection on Social Media Using Machine Learning Retraction Retraction : Fake News Detection on Social Media Using Machine Learning (J . Phys .: Conf . Ser . 1916 012235)'.

Shahbazi, Z. and Byun, Y.-C. (2021) 'Fake Media Detection Based on Natural Language Processing and Blockchain Approaches', *IEEE Access*, 9, pp. 128442–128453. Available at: https://doi.org/10.1109/ACCESS.2021.3112607.

Shaik, M.A. *et al.* (2023) 'Fake News Detection using NLP', in *International Conference on Innovative Data Communication Technologies and Application, ICIDCA 2023*

- *Proceedings*. Available at: https://doi.org/10.1109/ICIDCA56705.2023.10100305.

Shaikh, J. and Patil, R. (2020) ‘Fake news detection using machine learning’, in *Proceedings - 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security, iSSSC 2020*. Available at: https://doi.org/10.1109/iSSSC50941.2020.9358890.

Sharma, D.K. *et al.* (2023) ‘A Survey of Detection and Mitigation for Fake Images on Social Media Platforms’, *Applied Sciences*, 13(19). Available at: https://doi.org/10.3390/app131910980.

Sharma, K. *et al.* (2019) ‘Combating Fake News: A Survey on Identification and Mitigation Techniques’, *ACM Trans. Intell. Syst. Technol.*, 10(3). Available at: https://doi.org/10.1145/3305260.

Shoemaker, E. (2019) ‘Using data science to detect fake news’, p. 2010.

Shu, K. *et al.* (2017) ‘Fake News Detection on Social Media: A Data Mining Perspective’, *SIGKDD Explor. Newsl.*, 19(1), pp. 22–36. Available at: https://doi.org/10.1145/3137597.3137600.

Shu, K. *et al.* (2019) ‘Defend: Explainable fake news detection’, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 395–405. Available at: https://doi.org/10.1145/3292500.3330935.

Shu, K., Wang, S. and Liu, H. (2018) ‘Understanding User Profiles on Social Media for Fake News Detection’, *Proceedings - IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018*, pp. 430–435. Available at: https://doi.org/10.1109/MIPR.2018.00092.

Shu, K., Wang, S. and Liu, H. (2019) ‘Beyond news contents: The role of social context for fake news detection’, *WSDM 2019 - Proceedings of the 12th ACM*

International Conference on Web Search and Data Mining, (i), pp. 312–320. Available at: https://doi.org/10.1145/3289600.3290994.

Simonyan, K. and Zisserman, A. (2014) 'Very Deep Convolutional Networks for Large-Scale Image Recognition', *arXiv 1409.1556* [Preprint].

Sinclair, C. (2020) 'Parody: Fake News, Regeneration and Education', *Postdigital Science and Education*, 2(1), pp. 61–77. Available at: https://doi.org/10.1007/s42438-019-00054-x.

Singh, G. and Selva, K. (2023) 'A Comparative Study of Hybrid Machine Learning Approaches for Fake News Detection that Combine Multi-Stage Ensemble Learning and NLP-based Framework', *techrxiv* [Preprint]. Available at: https://doi.org/10.36227/techrxiv.21856671.v3.

Singhania, S., Fernandez, N. and Rao, S. (2017) '3HAN: A Deep Neural Network for Fake News Detection', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Available at: https://doi.org/10.1007/978-3-319-70096-0_59.

Siram Divya, S.D. (2022) 'Fake News Detection on Social Media usingDeep learning and Semantic Knowledge Sources', *Journal of Science and Technology*, 7(9), pp. 83–101. Available at: https://doi.org/10.46243/jst.2022.v7.i09.pp83-101.

Smajić, A., Grandits, M. and Ecker, G.F. (2022) 'Using Jupyter Notebooks for re-training machine learning models', *Journal of Cheminformatics* [Preprint]. Available at: https://doi.org/10.1186/s13321-022-00635-2.

Smitha, N. and Bharath, R. (2020) 'Performance Comparison of Machine Learning Classifiers for Fake News Detection', *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020*, 2020(December), pp. 696–700. Available at:

https://doi.org/10.1109/ICIRCA48905.2020.9183072.

Stets, J.E. and Burke, P.J. (2000) 'Identity theory and social identity theory', *Social Psychology Quarterly* [Preprint]. Available at: https://doi.org/10.2307/2695870.

Sundaram, J. *et al.* (2023) *An exploration of python libraries in machine learning models for data science, Advanced Interdisciplinary Applications of Machine Learning Python Libraries for Data Science*. Available at: https://doi.org/10.4018/978-1-6684-8696-2.ch001.

Tabassum, N. *et al.* (2024) 'Punny_Punctuators@DravidianLangTech-EACL2024: Transformer-based Approach for Detection and Classification of Fake News in Malayalam Social Media Text', *DravidianLangTech 2024 - 4th Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, Proceedings of the Workshop*, pp. 180–186.

Tandoc, E.C., Lim, Z.W. and Ling, R. (2018) 'Defining "Fake News": A typology of scholarly definitions', *Digital Journalism* [Preprint]. Available at: https://doi.org/10.1080/21670811.2017.1360143.

Taylor, G.W. *et al.* (2010) 'Convolutional Learning of Spatio-temporal Features', in *European Conference on Computer Vision*.

Thota, A. et al. (2018) 'Fake News Detection: A Deep Learning Approach', *SMU Data Science Review* [Preprint].

Tokpa, F.W.R. *et al.* (2023) 'Fake News Detection in Social Media: Hybrid Deep Learning Approaches', *Journal of Advances in Information Technology*, 14(3), pp. 606–615. Available at: https://doi.org/10.12720/JAIT.14.3.606-615.

Truică, C.O. and Apostol, E.S. (2023) 'It's All in the Embedding! Fake News Detection Using Document Embeddings', *Mathematics*, 11(3). Available at: https://doi.org/10.3390/math11030508.

Tschiatschek, S. *et al.* (2018) ‘Fake News Detection in Social Networks via Crowd Signals’, *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018,* pp. 517–524. Available at: https://doi.org/10.1145/3184558.3188722.

Tversky, A. and Kahneman, D. (1992) ‘Advances in prospect theory: Cumulative representation of uncertainty’, *Journal of Risk and Uncertainty*, 5(4), pp. 297–323. Available at: https://doi.org/10.1007/BF00122574.

Uma Sharma, Sidarth Saran, S.M.P. (2021) ‘Fake News Detection using Machine Learning Algorithms’, *International Journal of Engineering Research & Technology (IJERT),* 9(3), pp. 1–10. Available at: https://doi.org/10.1109/ICCAMS60113.2023.10525971.

Umer, M. *et al.* (2020) ‘Fake news stance detection using deep learning architecture (CNN-LSTM)’, *IEEE Access* [Preprint]. Available at: https://doi.org/10.1109/ACCESS.2020.3019735.

Upadhya, V. and Sastry, P.S. (2019) ‘An Overview of Restricted Boltzmann Machines’, *Journal of the Indian Institute of Science*, 99(2), pp. 225–236. Available at: https://doi.org/10.1007/s41745-019-0102-z.

Veeraiah, V. *et al.* (2024) ‘Fake News Detection using Natural Language Processing and TensorFlow in IoT System’, *ijisae*, 12(10), pp. 1–9.

Vereshchaka, A., Cosimini, S. and Dong, W. (2020) ‘Analyzing and distinguishing fake and real news to mitigate the problem of disinformation’, *Computational and Mathematical Organization Theory*, 26(3), pp. 350–364. Available at: https://doi.org/10.1007/s10588-020-09307-8.

Verma, A., Mittal, V. and Dawn, S. (2019) ‘FIND: Fake Information and News Detections using Deep Learning’, in *2019 12th International Conference on*

Contemporary Computing, IC3 2019. Available at: https://doi.org/10.1109/IC3.2019.8844892.

Vosoughi, S., Roy, D. and Aral, S. (2018) ‘The spread of true and false news online’, *Science* [Preprint]. Available at: https://doi.org/10.1126/science.aap9559.

Wang, A. (2010) ‘Don’t Follow Me - Spam Detection in Twitter.’, in, pp. 142–151. Available at: https://doi.org/10.7312/wang15140-003.

Wang, Z. (2022) ‘Deep Learning Methods for Fake News Detection’, in *2022 IEEE 2nd International Conference on Data Science and Computer Application, ICDSCA 2022*. Available at: https://doi.org/10.1109/ICDSCA56264.2022.9988556.

Wani, A. *et al.* (2021) ‘Evaluating Deep Learning Approaches for Covid19 Fake News Detection’, in *Communications in Computer and Information Science*. Available at: https://doi.org/10.1007/978-3-030-73696-5_15.

Waszak, P.M., Kasprzycka-Waszak, W. and Kubanek, A. (2018) ‘The spread of medical fake news in social media – The pilot quantitative study’, *Health Policy and Technology*, 7(2), pp. 115–118. Available at: https://doi.org/https://doi.org/10.1016/j.hlpt.2018.03.002.

Wixted, J.T. (2007) ‘Dual-process theory and signal-detection theory of recognition memory’, *Psychological Review* [Preprint]. Available at: https://doi.org/10.1037/0033-295X.114.1.152.

Wu, L. and Rao, Y. (2020) ‘Adaptive Interaction Fusion Networks for Fake News Detection’, *CoRR*, abs/2004.1.

Yadav, A.K. *et al.* (2023) ‘Fake News Detection Using Hybrid Deep Learning Method’, *SN Computer Science*, 4(6). Available at: https://doi.org/10.1007/s42979-023-02296-w.

Yazdi, K.M. *et al.* (2020) ‘Improving Fake News Detection Using K-means and Support

Vector Machine Approaches', *International Journal of Electrical, Electronic and Communication Sciences*, 13.0(2). Available at: https://doi.org/10.5281/zenodo.3669287.

Ye-Chan Ahn, C.-S.J. (2019) 'Natural Language Contents Evaluation System for Detecting Fake News using Deep Learning', *https://ieeexplore.ieee.org/xpl/conhome/8854238/proceeding*, pp. 1–23. Available at: https://doi.org/https://doi.org/10.1109/JCSSE.2019.8864171.

Yuffon, F. *et al.* (2020) 'Fake News Detection: A Machine Learning Approach using Automated-Text Analysis Technique', (October 2021), pp. 1–31.

Zhang, X. *et al.* (2016) 'Detecting Spam and Promoting Campaigns in Twitter', *ACM Trans. Web*, 10(1). Available at: https://doi.org/10.1145/2846102.

Zhang, X. and Ghorbani, A.A. (2020) 'An overview of online fake news: Characterization, detection, and discussion', *Information Processing & Management*, 57(2), p. 102025. Available at: https://doi.org/https://doi.org/10.1016/j.ipm.2019.03.004.

Zhou, T. *et al.* (2017) 'Unsupervised Learning of Depth and Ego-Motion from Video', in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 6612–6619. Available at: https://doi.org/10.1109/CVPR.2017.700.

Zhou, X. *et al.* (2019) 'Fake News: Fundamental Theories, Detection Strategies and Challenges', in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. New York, NY, USA: Association for Computing Machinery (WSDM '19), pp. 836–837. Available at: https://doi.org/10.1145/3289600.3291382.

Zubiaga, A. and Ji, H. (2013) 'Tweet, but Verify: Epistemic Study of Information Verification on Twitter', *CoRR*, abs/1312.5.

APPENDIX A:

PROPOSE CODE

```
# Importing libraries for data manipulation and analysis
#deals with structured data in DataFrames for pre-processing
import pandas as pnds
#supports various mathematical functions for numeric computations and
array manipulations
import numpy as nmpy
# Importing visualization libraries:
#statistical data visualisation
import seaborn as snb

#Word2Vec: to create word embeddings
from gensim.models import Word2Vec
# Wordcloud produces visual representation of word frequency
from wordcloud import WordCloud
# Importing data processing and model training utilities:
#for data balancing using resample
from sklearn.utils import resample

# Importing datasets as pandas DataFrames
#Fact-checked articles, fake news from GossipCop
GCF_Data = pnds.read_csv('/content/drive/MyDrive/ML PROJECTS/Kartik Fake
News/Dataset/gossipcop_fake.csv')
# Fact-checked articles, real news from GossipCop
GCR_Data = pnds.read_csv('/content/drive/MyDrive/ML PROJECTS/Kartik Fake
News/Dataset/gossipcop_real.csv')
# Fact-checked articles, fake news from PolitiFact
PFF_Data = pnds.read_csv('/content/drive/MyDrive/ML PROJECTS/Kartik Fake
News/Dataset/politifact_fake.csv')
# Fact-checked articles, real news from PolitiFact
PFR_Data = pnds.read_csv('/content/drive/MyDrive/ML PROJECTS/Kartik Fake
News/Dataset/politifact_real.csv')
# Imports CSV files from paths located in Google Drive, Every dataset
represents different fact-checking sources for articles.

# Labeling the data sets:
#(PolitiFact Fake) is tagged 0 (Fake news)
PFF_Data['label'] = 0
#(PolitiFact Real) labeled as 1 (Real news)
PFR_Data['label'] = 1
```

```
#column is added to every dataset with 0 for fake news and 1 for real
news.

# Join PolitiFact Fake and Real datasets into one unified DataFrame
PF_Combined = pnds.concat([PFF_Data, PFR_Data], ignore_index=True)
# The resulting DataFrame is shuffled randomly using sample(frac=1) so
that data is mixed
PF_Combined = PF_Combined.sample(frac=1).reset_index(drop=True)

# Use the `tokenize_text` function on the 'title' column of DataFrame
PF_Combined
PF_Combined['tokenized_title'] = PF_Combined['title'].apply(tokenize_text)
# - This would tokenize every title in the dataset into words.

# This allows us to check that the tokenization process was correct
print(PF_Combined[['title', 'tokenized_title']].head())

# Finding rows that have 'title' values duplicated in the DataFrame
PF_Combined
duplicate_titles = PF_Combined[PF_Combined.duplicated(subset=['title'],
keep=False)]
# The function `duplicated` checks for duplicates in column 'title'
# The resulting rows are put into the DataFrame `duplicate_titles`

# Print out how many rows have a title duplicated
print("\nNumber of duplicate titles:", len(duplicate_titles))

# Removing duplicate rows in PF_Combined based on 'title'
PF_Combined_no_duplicates = PF_Combined.drop_duplicates(subset=['title'],
keep='first')
# Dropping duplicate rows with 'title' as a reference column
# The 'keep' argument is set to 'first', such that only the first row of
the given duplicate 'title' will remain and all the others are dropped

# This shows a view of the data now that the rows with duplicates are
eliminated
print("\nDataFrame after removing duplicate titles:")
# Prints the number of rows left after removing duplicates
print(PF_Combined_no_duplicates.head())
# This is a form of double-checking what's left in the data after removing
duplicates
```

```
print("\nNumber of rows after removing duplicates:",
len(PF_Combined_no_duplicates))

PF_Combined_no_duplicates['tweet_count'] =
PF_Combined_no_duplicates['tweet_ids'].str.split(',').str.len()
PF_Combined_no_duplicates['word_count'] =
PF_Combined_no_duplicates['title'].apply(lambda x: len(str(x).split()))

fig21 = px.scatter(
    PF_Combined_no_duplicates,
    x='word_count',
    y='tweet_count',
    color='label',
    title='Tweet Count vs. Title Word Count by Label',
    labels={'word_count': 'Word Count in Title', 'tweet_count': 'Tweet
Count', 'label': 'Label'}
)
fig21.show()

PF_Combined_no_duplicates['title_length'] =
PF_Combined_no_duplicates['title'].str.len()

fig = px.box(PF_Combined_no_duplicates, x='label', y='title_length',
color='label',
             title='Distribution of Title Lengths by Label')
fig.show()

fig8 = px.scatter(
    PF_Combined_no_duplicates,
    x='title_length',
    y='tweet_count',
    size='label',
    color='label',
    title='Title Length vs. Tweet Count (Bubble Chart)',
    labels={'title_length': 'Title Length (characters)', 'tweet_count':
'Tweet Count'},
    size_max=20
)
fig8.show()

fig11 = px.density_heatmap(
    PF_Combined_no_duplicates,
```

```
    x='title_length',
    y='tweet_count',
    title='Heatmap of Title Length vs. Tweet Count',
    labels={'title_length': 'Title Length (characters)', 'tweet_count':
'Tweet Count'},
    color_continuous_scale='Viridis'
)
fig11.show()

fig22 = px.area(
    PF_Combined_no_duplicates,
    x='id',
    y='tweet_count',
    color='label',
    title='Tweet Count Across Records (Stacked by Label)',
    labels={'id': 'Record ID', 'tweet_count': 'Tweet Count', 'label':
'Label'}
)
fig22.show()

avg_word_count = PF_Combined_no_duplicates['word_count'].mean()
PF_Combined_no_duplicates['above_avg_word_count'] =
PF_Combined_no_duplicates['word_count'] > avg_word_count

fig23 = px.pie(
    PF_Combined_no_duplicates,
    names='above_avg_word_count',
    title='Proportion of Titles Above Average Word Count',
    labels={True: 'Above Average', False: 'Below Average'},
    color='above_avg_word_count',
    color_discrete_map={True: 'gold', False: 'purple'}
)
fig23.show()

PF_Combined_no_duplicates['avg_word_length'] =
PF_Combined_no_duplicates['title'].apply(lambda x: sum(len(word) for word
in str(x).split()) / len(str(x).split()) if x else 0)

fig15 = px.histogram(
    PF_Combined_no_duplicates,
    x='avg_word_length',
    nbins=10,
```

```
    title='Distribution of Average Word Length in Titles',
    labels={'avg_word_length': 'Average Word Length'},
    color_discrete_sequence=['orange']
)
fig15.show()

# Step 1: Separate the DataFrame into fake and real news by splitting
based on 'label' column.
fake_data = PF_Combined_no_duplicates[PF_Combined_no_duplicates['label']
== 0]
#all rows whose label is 0 which means it is a fake news all rows whose
label is 1, means it is real news
real_data = PF_Combined_no_duplicates[PF_Combined_no_duplicates['label']
== 1]

# Step 2: Class Balance using resampling of minority class.
if len(fake_data) > len(real_data):
  # If there are more false news than actual ones then resample the actual
news articles so they will equal the false ones
    real_data = resample(real_data, replace=True,
n_samples=len(fake_data), random_state=42)
    # If there are more actual news than false then resample the false
news to equal actual
else:
  # If there are more real news articles than the fake ones, resample the
fake news articles to equal the number of the real ones.
    fake_data = resample(fake_data, replace=True,
n_samples=len(real_data), random_state=42)

    # Extracting the embeddings from the final hidden state of BERT.
    embeddings = model_output.last_hidden_state
    # The attention ensures that only consider tokens not the padding
    attention_mask = encoded_input['attention_mask']
    #mask the embeddings by multiplying by the attention mask, thereby
ignoring padding tokens.
    masked_embeddings = embeddings * attention_mask.unsqueeze(-1)
    # Sum the embeddings along the token dimension
    summed = torch.sum(masked_embeddings, dim=1)
    # Rescale the added embeddings by using a count of non-padding words
to get an average.
    counts = torch.clamp(torch.sum(attention_mask, dim=1, keepdim=True),
min=1e-9)
```

```
    # Normalize the sum of embeddings by count of non-padding tokens, such
that an average would be obtained.
    mean_pooled_embedding = summed / counts

    # Step 8 Convert the tensor to a NumPy array and return the result
    return mean_pooled_embedding.numpy()[0]
#Call 'get_bert_embedding' for each title from DataFrame and save the
output back in the DataFrame
PF_Combined_no_duplicates['bert_embedding'] =
PF_Combined_no_duplicates['title'].apply(get_bert_embedding)
#Print out the first several rows of the DataFrame for your titles and the
related BERT embeddings.
print(PF_Combined_no_duplicates[['title', 'bert_embedding']].head())

# Get training results from trainer's log history
train_results = trainer.state.log_history
#is a list of dictionaries containing information regarding each training
step, losses

# Variable 'final_train_loss' to store final value of training loss
final_train_loss = None
# Iterate over 'train_results' to find entry where 'loss' key exists
for entry in train_results:
  # 'loss' key has the loss value at every training step
    if 'loss' in entry:
      #set to whatever value is found for the last loss in the log
history.
        final_train_loss = entry['loss']
# Print the final training loss value, with 4 decimal places displayed, to
indicate how the model performed on the train data at the end
print(f"\nFinal Training Loss: {final_train_loss:.4f}")

# Assign 0 in the column 'label' for all rows in the DataFrame GCF_Data
GCF_Data['label'] = 0
# For all rows in the DataFrame GCR_Data assign value 1 to column 'label
GCR_Data['label'] = 1

#combines both DataFrames into one ensures that the index is reset after
concatenation, producing a new sequential index
GC_Combined = pnds.concat([GCF_Data, GCR_Data], ignore_index=True)
#Resets the index following a shuffle, discarding the previous index
without adding it as a new column.
```

```
GC_Combined = GC_Combined.sample(frac=1).reset_index(drop=True)

fig8 = px.scatter(
    GC_Combined_no_duplicates,
    x='title_length',
    y='tweet_count',
    size='label',
    color='label',
    title='Title Length vs. Tweet Count (Bubble Chart)',
    labels={'title_length': 'Title Length (characters)', 'tweet_count':
'Tweet Count'},
    size_max=20
)
fig8.show()

fig17 = px.parallel_coordinates(
    GC_Combined_no_duplicates,
    dimensions=['title_length', 'tweet_count', 'label'],
    color='label',
    title='Parallel Coordinates: Title Length, Tweet Count, and Label',
    labels={'title_length': 'Title Length', 'tweet_count': 'Tweet Count',
'label': 'Label'},
    color_continuous_scale=px.colors.diverging.Tealrose
)
fig17.show()

fig22 = px.area(
    GC_Combined_no_duplicates,
    x='id',
    y='tweet_count',
    color='label',
    title='Tweet Count Across Records (Stacked by Label)',
    labels={'id': 'Record ID', 'tweet_count': 'Tweet Count', 'label':
'Label'}
)
fig22.show()

avg_word_count = GC_Combined_no_duplicates['word_count'].mean()
GC_Combined_no_duplicates['above_avg_word_count'] =
GC_Combined_no_duplicates['word_count'] > avg_word_count

fig23 = px.pie(
```

```
    GC_Combined_no_duplicates,
    names='above_avg_word_count',
    title='Proportion of Titles Above Average Word Count',
    labels={True: 'Above Average', False: 'Below Average'},
    color='above_avg_word_count',
    color_discrete_map={True: 'gold', False: 'purple'}
)

# Runs inference on the test dataset returning the predictions along with
probabilities and other details
predictions = nmpy.argmax(preds.predictions, axis=1)

# The raw predictions or logits, obtained from 'preds.predictions', need
to be transformed into class labels
print(classification_report(test_labels, predictions))

#Select the index corresponding to the maximum probability in each
prediction
cm = confusion_matrix(test_labels, predictions)
#will add values as annotations to the heatmap

# for displaying class names 'Fake' and 'Real'
# Now label axes and give a title to plot the confusion matrix
ptks.ylabel('Actual')
# 'ptks.title('Confusion Matrix')' the title for the plot
ptks.xlabel('Predicted')
# plots the confusion matrix
ptks.title('Confusion Matrix')
# print the plot
ptks.show()

def create_metric_graph(accuracy, precision, recall, f1):
  #creates a new figure using Plotly's graph object
    fig = go.Figure()
#creates a bar chart on the figure
# labels the x-axis with the metrics
    fig.add_trace(go.Bar(
        #represents the heights of the bars for each metric
        x=['Accuracy', 'Precision', 'Recall', 'F1-Score'],
        #displays metric values top of the bars with four decimal places
        y=[accuracy, precision, recall, f1],
        #automatically positions the text labels on the bars
```

```
        marker_color=['royalblue', 'lightcoral', 'lightgreen',
'lightsalmon'],
        #customizes a layout of the chart

#Get the training logs from the history of trainer's log
train_results = trainer.state.log_history

#contains log entries of the process of training: a list of loss values,
passed at different steps
final_train_loss = None
# Set ' final_train_loss' to None as this is where the last loss that is
recorded will be placed.
for entry in train_results:
  #Pass through the log entries looking for the last one with the loss
value.
    if 'loss' in entry:
      #if 'loss' in entry' checks whether the entry has a 'loss' key
        final_train_loss = entry['loss']
# Print the final training loss with decimal places
print(f"\nFinal Training Loss: {final_train_loss:.4f}")

#creates new figure to plot, size=8x6 inches
ptks.figure(figsize=(8, 6))
#KDE for titles
snb.kdeplot(PF_Combined_no_duplicates[PF_Combined_no_duplicates["label"]==
0]['title_length'], label="Fake", shade=True)
#takes the title length of the fake news, i.e. label == 0
snb.kdeplot(PF_Combined_no_duplicates[PF_Combined_no_duplicates["label"]==
1]['title_length'], label="Real", shade=True)
# The same steps are repeated for the real news (label 1), with the title
length data and the label "Real"
ptks.title('Distribution of Title Lengths for each Label')
#Title Lengths for each Label, adds a title to the plot
ptks.xlabel('Title Length')
#labels the x-axis, indicating the data represents the length of titles
ptks.ylabel('Density')
# displays the plot, showing the length of titles of fake and real news
ptks.show()
```