FAKE NEWS DETECTION SYSTEM USING MACHINELEARNING MODELS

By
Maria Iqbal

A thesis submitted to the faculty of Information Security Department, Military College of Signals, National University of Sciences and Technology, Islamabad, Pakistan, in partial fulfillment of the requirements for the degree of MS in Information Security

July 2023

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by **Maria Iqbal**, Registration No. **00000320998**, of **Military College of Signals** has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations/MS Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members and local evaluators of the scholar have also been incorporated in the said thesis.

Signature: ____________________

Name of Supervisor **Prof. Mian M Waseem Iqbal**

Date: 11/8/23

Signature (HOD): ____________________

Date: 11/8/23

HoD
Information Security
Military College of Sigs

Signature (Dean/Principal) ____________________

Date: 11/8/23

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

DEDICATION

I dedicate this thesis to myself, for achieving what once seemed impossible.

ACKNOWLEDGEMENTS

I am grateful to God Almighty who has bestowed me with the strength and the passion to accomplish this thesis and I am thankful to Him for His mercy and benevolence for without His consent I could not have completed this research. I would like to express my gratitude to all my family for their invaluable support and being my strength throughout in these very intense academic years.

Lastly, I would like to specially thank my supervisor Asst. Prof. Dr Mian Muhammad Waseem Iqbal for his constant guidance and immense support in shaping this thesis into a reality.

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

MLP	Multi-layer Perceptron
BiGRU	Bidirectional Gated Recurrent Unit.
ELMo	Embedding from Language Model
LDA	Latent Dirichlet Allocation
(RNN)	Recurrent Neural Network
GRU	Gated Recurrent Unit
NLP	Natural Language Processing
WHO	World Health Organization
CDC	Centers for Disease Control and Prevention
ML	Machine Learning
DL	Deep Learning
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
BERT	Bidirectional Encoder Representations from Transformers
RoBERTa	A variant of BERT model
XLNet	A generalized autoregressive pretraining model
LDA	Latent Dirichlet Allocation
GPT	Generative Pre-trained Transformer
RNN	Recurrent Neural Network
AUC	Area Under the Curve
F1-score	F1-measure, a metric for classification performance
Pr	Probability of being "real"

Pf	Probability of being "fake"
URL	Uniform Resource Locator (web address)
API	Application Programming Interface
POS	Part-of-Speech
NER	Named Entity Recognition
TF-IDF	Term Frequency-Inverse Document Frequency
ROC	Receiver Operating Characteristic

ABSTRACT

Social media's role in news consumption is a two-edged sword. On one hand, its accessibility, affordability, and swift information sharing encourage people to turn to social media for news. However, on the other hand, it also allows for the rampant spread of "fake news" – news with intentionally false information. The proliferation of fake news poses significant risks to individuals and society. Consequently, the detection of fake news on social media has become a prominent research area, garnering substantial attention.

Detecting fake news on social media indeed poses distinct challenges and exhibits unique characteristics that make traditional detection algorithms inadequate or unsuitable for this task. Firstly, fake news is deliberately crafted to deceive readers with false information, making its detection based solely on news content difficult and complex. Hence, auxiliary information, such as users' social engagements on social media, becomes crucial in making accurate determinations. Secondly, leveraging this auxiliary information poses its own set of challenges, as users' interactions with fake news generate large, incomplete, unstructured, and noisy data.

Given the challenging and relevant nature of fake news detection on social media, to facilitate additional investigation into this issue, we ran a survey. The survey provides a comprehensive review of detecting fake news on social media, covering various aspects such as fake news characteristics based on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics, and representative datasets. We also discuss related fields of study, unresolved issues, and future research paths for identifying bogus news on social media.

Chapter 1 INTRODUCTION

Fake news encompasses fabricated information and deceptive content disseminated through traditional platforms and online channels, notably social media. In recent times, there has been a growing focus on fake news within the realm of social media, driven by the prevailing political climate and other contributing factors. The detection of misinformation on social media is both crucial and technologically demanding. This challenge arises in part because even humans struggle to accurately discern between false and true news, as it involves laborious evidence gathering and meticulous fact-checking. Given the growth of skill and the escalating proliferation of deceptive articles on social media, the development of automated frameworks for identifying fake news has become increasingly imperative.

This paper presents the system, which conducts binary classification on social media tweets to categorize them as either "real" or "fake". I am following the “A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection” paper to verify that the mentioned things or accuracy is according to the paper or not. For this reason, I updated the dataset took different data of COVID-19 and applied all of the models existing models on that. To enhance the efficiency of this approach, I have utilized transfer learning, a technique that has demonstrated significant effectiveness in text classification tasks. Since each model doesn't need to be trained from scratch with this method, training time is reduced. Text preprocessing, model prediction, tokenization and ensemble edifice utilizing a soft voting technique are the main steps in our methodology. Later analysis, we added a heuristic post-processing technique to our false news detection engine that takes into account crucial tweet components like tweets handles and labels. This methodology had resulted in markedly greater results compared to the leading entry on the official leaderboard. Additionally, they have included examples of tweets in which the post-processing method correctly predicted the outcome compare the original categorization output showcasing its effectiveness. I have used the same models same approach but with new data set and results was totally different.

1.1.Motivation

The motivation behind reconstructing the models and evaluating their results using a new dataset in the research paper titled "A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection" stems from the need to enhance the existing methods for detecting fake news specifically related to the COVID-19 pandemic. As the COVID-19 crisis unfolded, there was a surge in misinformation and spreading on social media networks is false information, which had severe consequences on public health and safety. Therefore, it became crucial to develop more effective and reliable techniques to combat the spread of fake news surrounding COVID-19.

By reconstructing the models and utilizing a new dataset, the researchers aimed to improve upon previous approaches and address the unique challenges posed by COVID-19-related misinformation. The creation of an ensemble framework driven by heuristics allowed for the integration of heuristic rules and techniques to better detention the distinguishing features of fake news related to the pandemic. This approach takes into account various aspects such as linguistic patterns, contextual information, and the source credibility to make more accurate determinations.

The evaluation of the reconstructed models and their results using the new dataset provides valuable insights into the performance and efficacy of the proposed framework. It allows researchers and practitioners to evaluate the strategy's efficiency in accurately detecting and classifying COVID-19-related fake news. By comparing the results with existing methods and potentially benchmarking against other approaches, the research paper contributes to the ongoing efforts in combating misinformation and enhancing the reliability of information during the COVID-19 pandemic.

1.2.Problem Statement

The purpose of introducing or researching on fake news detection models are:

1. **Impact on Public Opinion:** Fake news has the potential to influence public opinion and decision-making processes, leading to distorted perceptions and misinformation-driven actions. Therefore, developing fake news detection models can help mitigate the

negative impact on public opinion and ensure informed decision-making.

2. **Threat to Democracy:** The dissemination of fake news poses a significant threat to democratic processes by manipulating public discourse, election campaigns, and political debates. Creating effective fake news detection models can help safeguard the integrity of democratic systems and protect the public from manipulation and misinformation.
3. **Social Division and Polarization:** Fake news often contributes to social division and polarization by spreading biased or inflammatory narratives. Developing accurate detection models can aid in identifying and countering fake news that fuels societal divisions and promotes hostility.
4. **Economic Consequences:** Fake news can harm businesses and economies by spreading false information that impacts consumer behavior, stock markets, and investor confidence. Implementing robust fake news detection models can help mitigate the economic consequences of misinformation and protect financial stability.
5. **Public Safety and Health:** Misinformation regarding public safety measures, health-related issues, or medical treatments can have severe consequences on public health and safety. By detecting and debunking fake news, models can help ensure accurate information dissemination and promote public well-being.
6. **Trust in Media:** The proliferation of fake news erodes trust in media organizations, leading to skepticism and disbelief in legitimate news sources. Developing reliable fake news detection models can help restore trust by enabling the identification and separation of credible news from false or misleading information.
7. **Online Security:** Fake news is often propagated through online platforms, which can be exploited by malicious actors for various purposes, including phishing attacks, identity theft, or the spread of malware. Creating effective detection models can contribute to enhancing online security and protecting users from such cyber threats.
8. **Legal and Ethical Implications**: The spread of fake news can have legal and ethical implications, such as defamation, infringement of intellectual property rights, or privacy violations. Developing accurate detection models can assist in identifying and addressing such violations, promoting legal compliance and ethical practices.

9. **Algorithmic Bias and Fairness:** Fake news detection models need to be developed with attention to potential biases, ensuring fairness in their outcomes and avoiding amplification of existing prejudices. Addressing algorithmic bias and fairness concerns is essential to build trust and ensure equitable treatment across diverse user groups.

10. **International Disinformation Campaigns:** Fake news is often used as a tool for disinformation campaigns by state actors or foreign entities to manipulate public opinion, sow discord, or interfere in political processes. Developing effective detection models can aid in identifying and countering these disinformation campaigns, protecting the sovereignty and integrity of nations.

1.3.Research Objectives

The main objectives of research on fake news detection systems are:

1. **Accuracy:** Developing accurate and reliable algorithms that can effectively detect and classify fake news from genuine information. The primary goal is to minimize false positives and false negatives, ensuring that the detection system can reliably differentiate between real and fake news.

2. **Scalability:** Creating scalable and efficient models that can handle the vast amount of data generated on social media platforms. The objective is to build detection systems that can process and analyze large volumes of information in real-time, considering the dynamic nature of news dissemination on social media.

3. **Robustness:** Designing detection systems that are resilient to various tactics used to deceive users, such as misleading headlines, manipulated images, or deceptive content. The objective is to develop models that can adapt to evolving techniques employed by creators of fake news, making the detection system more robust and effective.

4. **Multilingual and Multimodal Support:** Extending the detection systems to handle different languages and diverse types of media content, including text, images, and videos. The objective is to ensure that the detection models can identify fake news across different languages and media formats, catering to the global nature of social media platforms.

5. **Real-time Detection:** Enabling real-time detection and classification of fake news, ensuring timely responses to mitigate the spread of misinformation. The objective is to develop systems that can identify and flag fake news as quickly as possible, allowing for prompt interventions and fact-checking efforts.

6. **Explainability:** Enhancing the transparency and interpretability of fake news

detection systems by providing explanations or evidence for the classification decisions. The objective is to build models that can clarify why a portion of news is classified as fake, enabling users and stakeholders to understand the basis for the system's output.

7. **User Empowerment:** Developing tools and systems that empower users to identify and evaluate the credibility of news themselves. The objective is to provide users with the necessary information, resources, and critical thinking skills to assess the authenticity and reliability of news sources.
8. **Collaboration and Data Sharing:** Promoting collaboration among researchers, organizations, and social media platforms to share data, methodologies, and insights to collectively combat fake news. The objective is to foster a collaborative environment that facilitates the development of robust and effective detection systems through shared knowledge and resources.
9. **Ethical Considerations:** Incorporating ethical considerations into the design and deployment of fake news detection systems, such as ensuring fairness, minimizing biases, and respecting privacy rights. The objective is to develop systems that are ethically sound and align with societal values and norms.
10. **Real-world Impact:** Ensuring that the research on fake news detection systems translates into practical applications and has a positive impact on society, by reducing the spread of misinformation, fostering media literacy, and promoting a more informed public discourse.

1.4. Contributions

I am attempting to recalculate the results using a different dataset in fake news detection models, the potential contributions could include:

1. **Generalizability:** By using a different dataset, the research contributes to assessing the generalizability and robustness of the fake news detection models beyond the original dataset. It helps evaluate the performance and effectiveness of the models across diverse sources of fake news, ensuring that the findings are not limited to a specific dataset or context.
2. **Comparative Analysis:** Comparing the results obtained from different datasets

allows for a comparative analysis of the performance of fake news detection models. It helps identify patterns, strengths, and weaknesses of the models in different data environments, providing insights into the reliability and transferability of the detection techniques.

3. **Validation of Findings:** Replicating and recalculating results using a different dataset contributes to the validation of the initial findings. It helps confirm the robustness and consistency of the models' performance and reinforces the reliability of the original research outcomes.
4. **Dataset Bias Analysis:** Assessing the presentation of fake news recognition models on a new dataset allows for a closer examination of potential biases present in the models. It helps identify whether the models exhibit biases specific to the original dataset or if the performance is consistent across different datasets. This analysis contributes to addressing algorithmic fairness and bias concerns in fake news detection.
5. **Real-World Applicability:**

   Using a different dataset enhances the applicability of the fake news detection models to real-world scenarios. It aids in assessing how well the models perform in identifying false news across a variety of domains, contexts, or specific events, contributing to their practical utility and deployment in different settings.
6. **Algorithm Comparison:** Recalculating the results using a different dataset allows for a direct comparison of different fake news detection algorithms or techniques. It enables researchers to identify which methods perform better or worse on the new dataset, aiding in the selection and refinement of detection approaches.
7. **Novel Insights:** Exploring a different dataset in fake news detection models can deliver new perceptions into the nature of mis-information, its characteristics, and the challenges associated with detection. It may reveal unique patterns, features, or dynamics specific to the new dataset, leading to novel research directions and advancements in the field.
8. **Improvement Opportunities:** Analyzing the results obtained from a different dataset may uncover areas for improvement in the fake news detection models. It helps identify specific challenges or limitations that were not apparent in the original dataset, guiding future research and development efforts to enhance the accuracy and

effectiveness of the models.

Overall, recalculating the results using a different dataset contributes to the broader understanding of mis-information recognition, strengthens the validity of the research findings, and provides insights into the performance and applicability of the models in varied contexts and scenarios.

1.5. Thesis Outline

The following chapters comprise the organization and distribution of the research:

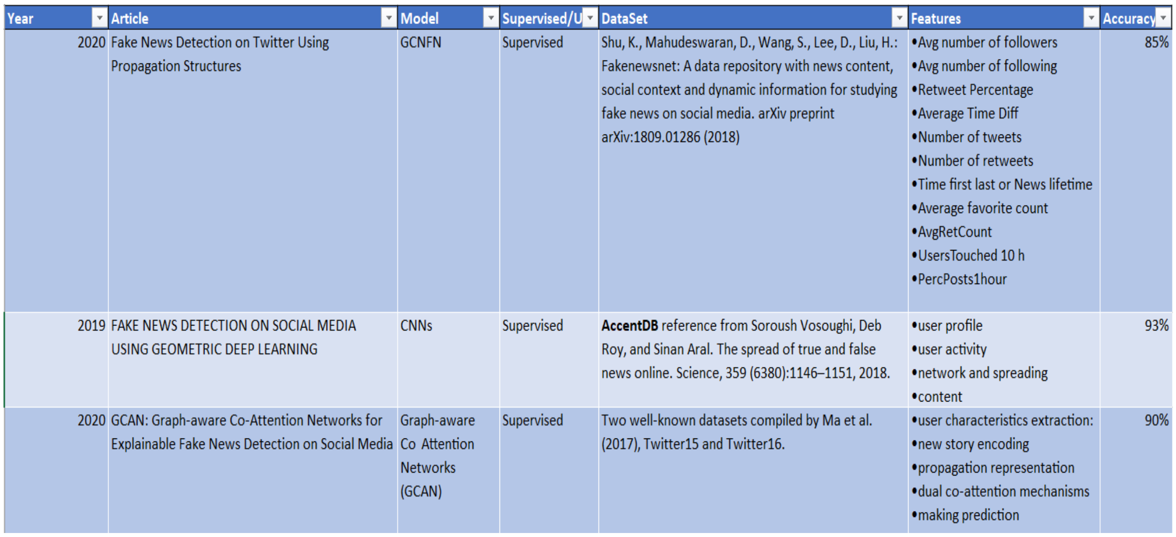
- **Chapter 1:** After a brief introduction, the problem statement and the reason for the research are underlined. There are a number of research goals. The contributions made as a result of this research are also recognized.
- **Chapter 2:** Presents an overview of other researches have been done on this topic. Also describe about the fake news detection and language models used in fake news detection mechanism.
- **Chapter 3:** This chapter summarizes the research implementation and results we get as a part of research.
- **Chapter 4:** Methodology is a crucial section of a research paper that outlines the systematic approach used to conduct the study or experiment. It provides a detailed description of the procedures, techniques, tools, and data exploration approaches engaged to answer the research queries or achieve the research objectives.
- **Chapter 5:** The dataset used in our research is a comprehensive collection of text data aimed at facilitating the detection and analysis of fake news. The dataset is specifically curated to address the challenges posed by misinformation, particularly in the context of COVID-19 news.
- **Chapter 6:** This chapter summarizes the research with conclusion drawn and delivers directions for upcoming work.

Chapter 2 Background and Related work

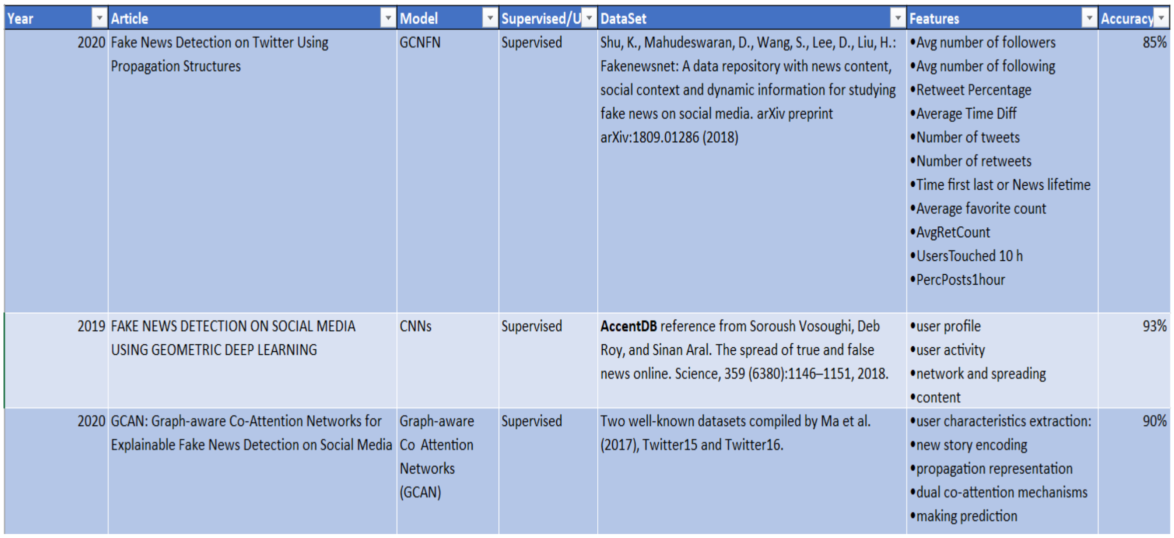
2.1 Background

Numerous researchers have contributed to tackling the challenge of automatic fake news detection through innovative approaches. For instance, one group of [1] this study proposes a fake news detection method that leverages temporal patterns of social context. The authors utilize a deep learning model to capture sequential patterns of user interactions and social dynamics to identify fake news.

Another research team [2] identified three key characteristics of fake news articles: textual data of the article, user response, and the promotion by source users. They proposed the CSI model, comprising Detention, Integrate modules and Score. The initial module employed Recurrent Neural Network (RNN) to detention the temporal representations of articles, while the subsequent module focused on user performance. The last module integrated the outputs from the initial models to identify fake news articles. Additionally, some earlier studies [3] leveraged news content with social perspective information to build fake news detection models, and [4] incorporated speaker profiles into an LSTM-grounded hybrid model for detecting mis-information in China, improving accuracy.



Year	Article	Model	Supervised/U	DataSet	Features	Accuracy
2020	Fake News Detection on Twitter Using Propagation Structures	GCNFN	Supervised	Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:1809.01286 (2018)	•Avg number of followers •Avg number of following •Retweet Percentage •Average Time Diff •Number of tweets •Number of retweets •Time first last or News lifetime •Average favorite count •AvgRetCount •UsersTouched 10 h •PercPosts1hour	85%
2019	FAKE NEWS DETECTION ON SOCIAL MEDIA USING GEOMETRIC DEEP LEARNING	CNNs	Supervised	**AccentDB** reference from Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. Science, 359 (6380):1146–1151, 2018.	•user profile •user activity •network and spreading •content	93%
2020	GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media	Graph-aware Co Attention Networks (GCAN)	Supervised	Two well-known datasets compiled by Ma et al. (2017), Twitter15 and Twitter16.	•user characteristics extraction: •new story encoding •propagation representation •dual co-attention mechanisms •making prediction	90%



Year	Article	Model	Supervised/U	DataSet	Features	Accuracy
2020	Fake News Detection on Twitter Using Propagation Structures	GCNFN	Supervised	Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:1809.01286 (2018)	•Avg number of followers •Avg number of following •Retweet Percentage •Average Time Diff •Number of tweets •Number of retweets •Time first last or News lifetime •Average favorite count •AvgRetCount •UsersTouched 10 h •PercPosts1hour	85%
2019	FAKE NEWS DETECTION ON SOCIAL MEDIA USING GEOMETRIC DEEP LEARNING	CNNs	Supervised	**AccentDB** reference from Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. Science, 359 (6380):1146–1151, 2018.	•user profile •user activity •network and spreading •content	93%
2020	GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media	Graph-aware Co Attention Networks (GCAN)	Supervised	Two well-known datasets compiled by Ma et al. (2017), Twitter15 and Twitter16.	•user characteristics extraction: •new story encoding •propagation representation •dual co-attention mechanisms •making prediction	90%

Another study [5] developed a model based on linguistic surface-level patterns using the LIAR dataset. Convolutional neural networks, logistic regression, support vector machines, and long short-term memory networks were among the reference point models. Additionally, they created a revolutionary hybrid convolutional neural network that included text and metadata, greatly enhancing the identification of bogus news at the finer granularity levels.

In the Shared Task, a different team of academics [6] provided a reliable and uncomplicated strategy for identifying fake news spreaders. Semantics, word classes, and other basic features were used in their approach, which was then used to classify data using a Random Forest model. In a different study, an innovative method for detecting disinformation on social media platforms was developed using news broadcast channels with recurrent and convolutional networks to collect both global and local user features [7].

A novel set of features that were taken from news content, the news basis, and the surroundings were provided in a recent study [8]. They used a variety of traditional and cutting-edge classifiers, such as Naive Bayes, k-Nearest Neighbors, Support Vector Machine, Random Forest with XGBoost, and RBF kernel, to calculate the prediction accuracy of existing techniques and features for automatic false news detection.

Another recent study [9] compared two end-to-end deep neural architecture variations for detecting fake news across multi-domain platforms. Encoding Layer (Bi-GRU), Embedding Layer, Multi-layer Perceptron (MLP), and Word-level Attention were all components of the initial model, which was grounded on Bidirectional Gated Recurrent Unit (BiGRU). The MLP Network and Embedding from Language Model (ELMo) were the foundations of the second model.

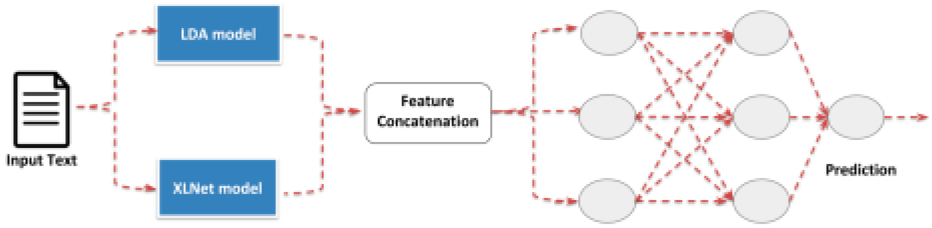


Figure 2:1 The recommended model structural

Fig. 1: The recommended model structural design combines XLNet, a contextualized language model, with Latent Dirichlet Allocation (LDA), a topic modeling technique. The goal is to leverage both contextualized representations and topic embeddings for disinformation detection.

First, the XLNet model is used to obtain contextualized representations of the input text. XLNet can capture the context and meaning of words based on their surrounding context in a sentence or document. Next, the LDA model is employed to obtain topic embeddings from the text. LDA is a probabilistic model that identifies latent topics present in a collection of papers. Each paper is characterized as a combination of topics, and every topic is represented as a distribution of words. The contextualized representations and topic embeddings are then concatenated to create a fused representation of the input text, capturing both the context and the underlying topics.

This fused demonstration is delivered through two fully coupled layers, which can learn complex patterns and relationships in the data. The amount produced of the last fully connected layer is then fed to a Softmax Layer, which implements the job of fake news detection by predicting the probability of the input text belonging to either the "fake" or "real" class.

In summary, the proposed model combines XLNet's contextualized representations with LDA's topic embeddings to create a comprehensive representation of the input text, which is then used for fake news detection using a combination of fully connected layers and a Softmax Layer.

There was another research group [10] compared the performance of the BERT language model against traditional machine learning methodologies for fake news recognition. BERT's ability to capture contextual information significantly improves detection accuracy. In [11] another paper applies geometric deep learning techniques to notice disinformation in social media. The approach represents social media data as graphs and employs graph convolutional networks for detection.

2.1.1 Fake News Detection

Fake News Detection is the process of using various techniques, algorithms, and machine learning models to identify and distinguish between genuine and deceptive news articles or information. The benefit of fake news bulletin detection is to combat the spread of misinformation and disinformation, thereby promoting the dissemination of accurate and reliable information to the public.

Disinformation has been in existence for centuries, emerging alongside the widespread circulation of news after the origination of the printing press in 1439. Conversely, there is no universally agreed-upon definition of "fake news." Hence, we begin by discussing and comparing various definitions of fake news found in the current collected works. Subsequently, we present our own definition of fake news.

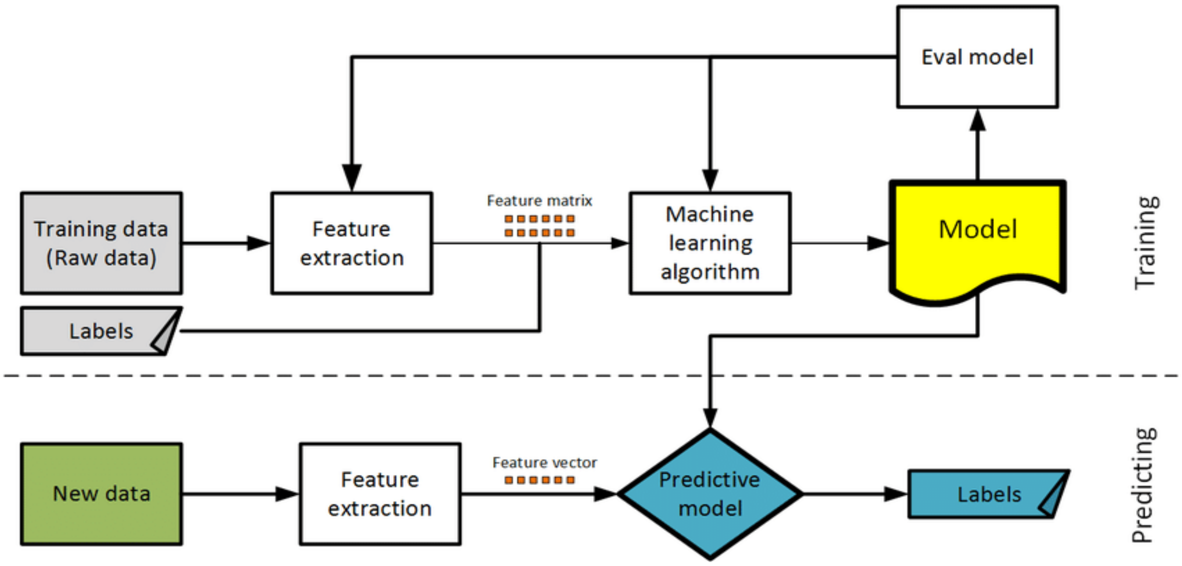


Figure 2:2 -Fake news detection process

As mentioned in above fig 3 for this process we gather data from various sources, including news websites, social media, and fact-checking organizations, to create a diverse dataset holding both fake and real news articles. Clean and preprocess the collected data by discarding noise, special characters, URLs, and irrelevant information to prepare it for analysis. After that extract relevant features from the text data to represent the articles effectively. Features may include n-grams, linguistic properties, sentiment analysis, and other textual characteristics. Once characteristics are chosen, a variety of machine learning algorithms, including support vector machines, LSTM, random forests, logistic regression, and transformers, are used to construct prediction models for false news identification. Combine the outputs of multiple models using ensemble techniques like soft voting, stacking, or weighted averaging to improve detection accuracy and robustness. Then measure the performance of the false news identification models using evaluation metrics such as precision, accuracy, F1-score, recall, and AUC-ROC to assess their effectiveness. Deploy the trained fake news detection model to analyze news articles in simultaneously and continuously monitor the system's performance to adapt to evolving fake news strategies.

1 Challenges in Fake News Detection:

- Misleading Language: false news articles often employ persuasive language and misleading headlines to deceive readers.
- Evolving Tactics: Adversaries continuously adapt their strategies to evade detection, making it challenging to keep up with new forms of deception.
- Social Context: The spread of fake news is influenced by user behavior, echo chambers, and confirmation bias, which adds complexity to the detection process.
- Satirical Content: Satire and parody articles may contain false information, blurring the lines between genuine news and fake news.

2 Importance of Fake News Detection:

False news identification plays a critical role in ensuring the reliability and trustworthiness of the information disseminated through various media channels. By distinguishing between authentic and deceptive content, it helps safeguard public discourse, enhance media literacy, and promote a more informed society. Effective fake news detection can contribute to reducing the harmful effects of misinformation, improving decision-making processes, and upholding the integrity of journalism in the digital age.

2.1.2 Chronological review:

The linear review provides an overview of the research contributions made in the field of detecting fake news using deep learning and machine learning-based algorithms. The representation in Fig. 2 illustrates the distribution of these contributions over the years. In 2018, 1.5% of the research works were published, followed by 9.2% in 2019, 30.7% in 2020, and the highest proportion of 58.4% in 2021. This trend indicates a growing interest and focus on this area, encouraging researchers to explore innovative techniques in the coming years.

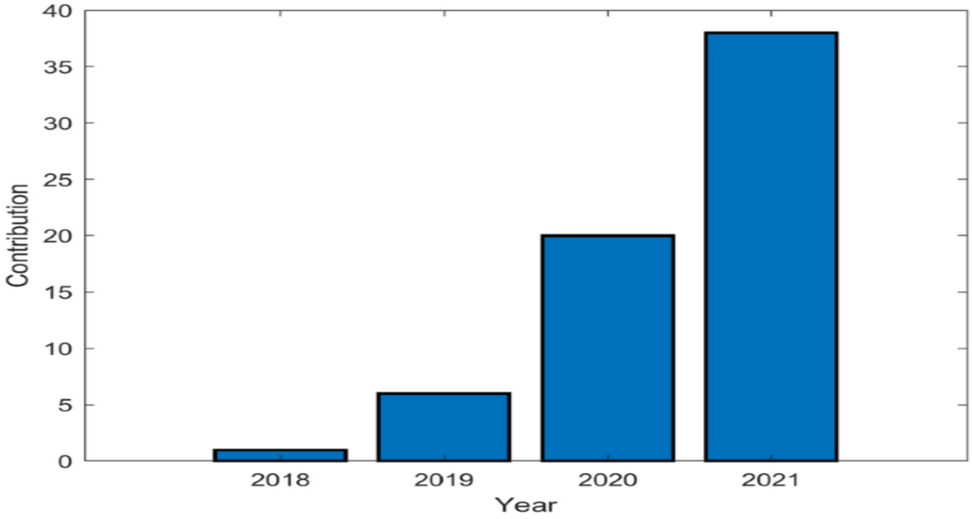


Figure 2:3- A chronological review of the existing fake news detection models

2.2 Fake News Characterizations

In this segment, we delve into the fundamental psychological theories and social pertaining to false news while exploring more sophisticated forms that emerge in the context of social media. Initially, we present a comprehensive examination of the diverse definitions of fake news, drawing distinctions between related concepts that are frequently misconstrued as fake news. Subsequently, we explore the different facets of fake news as observed in traditional media and contrast them with novel patterns that have emerged on social media platforms.

2.2.1Definitions of Fake News

Fake news has been present for an extensive period, dating back to the widespread circulation of news after the invention of the printing press in 14397. Despite its historical existence, there remains a lack of consensus on the precise meaning of "fake news." Thus, here, we aim to address this issue by examining and comparing several commonly used definitions of fake news found in paper. Subsequently, we present our own adopted definition, which will be utilized throughout this study.

A narrower explanation of false news refers to news articles deliberately and verifiably disseminating false information with the intent to mislead readers [2]. This definition highlights two critical features: the authenticity of the information, which can be objectively verified as false, and the intention of the creators to deceive consumers. Many recent studies have embraced this particular definition [57, 17, 62, 41].

Alternatively, wider definitions of false news revolve around either the genuineness or intent of the news content. For instance, some papers consider satire news as false news due to its false nature, even though satire is often meant for entertainment and is overt about its deceptive intent for consumers [9,67, 37, 4]. Other work directly includes misleading news within the realm of false news [66], encompassing serious fabrications, hoaxes, and satirical content.

For the purposes of this artifact, we adopt the narrower definition of false news. Officially, we state this definition as follows:

Definition 1 (Fake News) Fake news is a news article that is intentionally and verifiably false.

There are three main motives for selecting this narrow definition. Initially, focusing on the intent of fake news offers both practical and theoretical advantages, enabling a more profound understanding and analysis of the subject. Secondly, any truth verification methods applicable to the narrow definition can also be employed in the broader context. Lastly, this definition helps to clarify distinctions between false news and linked concepts not covered in this article. Ideas such as satire news with proper context, rumors unrelated to news events, difficult-to-verify conspiracy theories, unintentional misinformation, and hoaxes intended for fun or scams do not fall under our definition of fake news.

2.2.2Fake News on Traditional News Media

Fake news is not a novel issue, and its presence has evolved within the media landscape over the years. The shift from newsprint to television/radio and, more recently, to online news and social media platforms has significantly impacted the ecology of fake news. We refer to the period before social media significantly influenced the production and spread of fake news as "traditional fake news."

In the following sections, we will explore various social science and psychological principles that elucidate the influence of false news on both individuals and the wider social data network. These foundations provide valuable insights into understanding the effects and consequences of fake news at both micro and macro levels.

2.2.3 Fake News on Social Media

In this section, we'll debate a few distinctive appearances of bogus news on social media. We'll pay close attention to the main traits of false news that social media makes possible. Because social media shares many traits with conventional fake news, keep in mind that it can be used to spread false information as well. social media accounts that spread propaganda and are malicious. Despite the fact that most social media users are good, some of them could be false or even harmful. The affordability of creating social media accounts has led to the emergence of malicious users, including cyborg users, trolls and social bots. Social bots, handled by computer algorithms, are designed to mechanically produce content and relate with users on social media. Some social bots have malicious intent, such as spreading fake news, which was evident during the 2016 U.S. presidential election, where they distorted online debates on a large scale. Trolls are actual people who use the internet with the intention of upsetting online communities and evoking strong emotions in others. They play a significant role in the dissemination of fake news, and evidence suggests that paid Russian trolls were involved in spreading fake news during the election. Trolling behavior thrives in online discussions, influencing people's emotions and facilitating the spread of fake news by triggering negative emotions like anger and fear.

Users who are cyborgs combine automated tasks with human input. Although these profiles were created by people, the social media tasks are carried out by automated systems. The ability of cyborg users to flip between human and bot functionalities gives them special opportunity to disseminate false information.

One major challenge in combating social media's echo chamber impact is bogus news. Social media alters the way users seek and consume information, moving from a mediated form to a more disinter-mediated approach. Users are exposed to content that aligns with their existing beliefs, creating groups of like-minded individuals and reinforcing their opinions. The echo chamber effect fosters a sense of credibility based on others' perceptions and a preference for frequently encountered information, even if it is bogus news. Consequently, users tend to consume and believe bogus news due to limited exposure to diverse perspectives.

Homogeneous communities in echo chambers become the most important drivers of reinforcing polarization, information diffusion and further limiting the information ecosystem. Research has shown that exposure to repeated information increases positive opinions of that information, contributing to the reinforcement of fake news beliefs within these segmented communities.

Overall, the proliferation of fake news on social media is fueled by malicious users, the echo chamber effect, and the preference for like-minded information consumption, presenting significant challenges for dispelling misinformation and promoting a more diverse and accurate information environment.

2.3 Fundamental Theories

Fundamental human cognition and behavior theories developed across various disciplines, such as social sciences and economics, provide invaluable insights for fake news analysis. These theories can introduce new opportunities for qualitative and quantitative studies of big fake news data [Zhou et al. 2019a]. These theories can also facilitate building well-justified and explainable models for fake news detection and intervention, which, to date, have been rarely available [Miller et al. 2017]. We have conducted a comprehensive literature survey across various disciplines and have identified well-known theories that can be potentially used to study fake news. These theories are provided in Table 2 along with short descriptions, which are related to either (I) the news itself or (II) its spreaders. I. News-related theories. News-related theories reveal the possible characteristics of fake news content compared to true news content. For instance, theories have implied that fake news potentially differs from the truth in terms of, e.g., writing style and quality (by Undeutsch hypothesis) [Undeutsch 1967], quantity such as word counts (by information manipulation theory) [McCornack et al. 2014], and sentiments expressed (by four-factor theory) [Zuckerman et al. 1981]. It should be noted that these theories, developed by forensic psychology, target deceptive statements or testimonies (i.e., disinformation) but not fake news, though these are similar concepts. Thus, one research opportunity is to verify whether these attributes (e.g., information sentiment polarity) are statistically distinguishable among disinformation, fake news, and the truth, in particular, using big fake news data. On the other hand, these (discriminative) attributes identified can be used to automatically detect fake news using its writing style, where a typical study using supervised learning can be seen in [Zhou et al. 2019a]; we will provide further details in Section 3. II. User-related theories. User-related theories investigate the characteristics of users involved in fake news activities, e.g., posting, forwarding, liking, and commenting. Fake news, unlike information such as fake reviews [Jindal and Liu 2008], can "attract" both malicious and normal users [Shao et al. 2018]. Malicious users (e.g., some social bots [Ferrara et al. 2016]) spread fake news often intentionally and are driven by benefits [Hovland et al. 1957; Kahneman and Tversky 2013]. Some normal users (which we denote as vulnerable normal users) can frequently and unintentionally spread fake news without recognizing the falsehood. Such vulnerability psychologically stems from (i) social impacts and (ii) self-impact, where theories have been accordingly categorized and detailed in Table 2. Specifically, as indicated by the bandwagon effect [Leibenstein 1950], normative influence theory [Deutsch and Gerard 1955], social identity theory [Ashforth and Mael 1989], and availability cascade [Kuran and Sunstein 1999], to be liked and/or accepted by the community, normal users

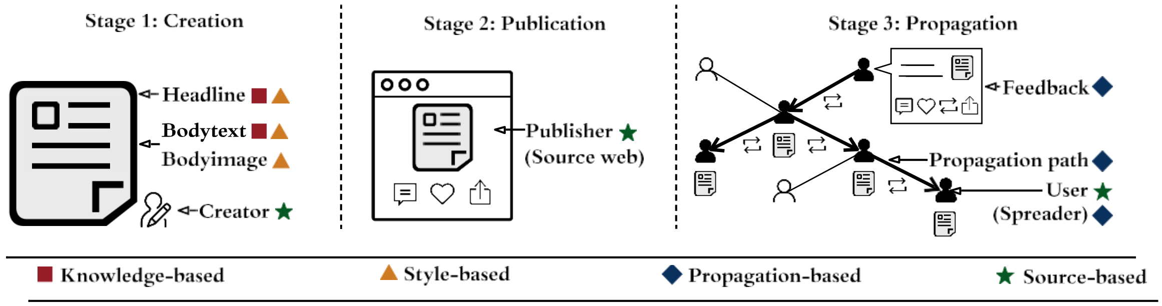


Figure 2:4- Fake News Life Cycle and Connections to the Four Fake News Detection Perspectives Presented

are encouraged to engage in fake news activities when many users have done so (i.e., peer pressure). One's trust to fake news and his or her unintentional spreading can be promoted as well when being exposed more to fake news (i.e., validity effect) [Boehm 1994], which often takes place due to the echo chamber effect on social media [Jamieson and Cappella 2008]. Such trust to fake news can be built when the fake news confirms one's preexisting attitudes, beliefs or hypotheses (i.e., confirmation bias [Nickerson 1998], selective exposure [Freedman and Sears 1965], and desirability bias [Fisher 1993]), which are often perceived to surpass that of others [Dunning et al. 1990; Pronin et al. 2001; Ward et al. 1997] and tend to be insufficiently revised when new refuting evidence is presented [Bálint and Bálint 2009; Basu 1997]. In such settings, strategies for intervening fake news from a user perspective (more discussions on fake news intervention are in Section 6) should be cautiously designed for users with different levels of credibility or intentions, even though they might all engage in the same fake news activity. For instance, it is reasonable to intervene with the spread of fake news by penalizing (e.g., removing) malicious users, but not for normal accounts. Instead, education and personal recommendations of true news articles and refuted fake ones can be helpful for normal users [Vo and Lee 2018]. Such recommendations should not only cater to the topics that the users want to read but should also capture topics that users are most gullible to. In Section 5, we will provide the path for utilizing these theories, i.e., quantifying social and self-impact, to enhance fake news research by identifying user intent and evaluating user credibility. Meanwhile, we should point out that clearly understanding the potential roles that the fundamental theories listed in Table 2 can play in fake news research requires further in-depth investigations of interdisciplinary nature.

Table 2-1: Fundamental theory of Fake News

	Theory		**Phenomenon**
News-related Theories	*Undeutsch hypothesis* [Undeutsch 1967]		A statement based on a factual experience differs in content style and quality from that of fantasy.
	Reality monitoring [Johnson and Raye 1981]		Actual events are characterized by higher levels of sensory- perceptual information.
	Four-factor theory [Zuckerman et al. 1981]		Lies are expressed differently in terms of arousal, behavior control, emotion, and thinking from truth.
	Information manipulation theory [McCornack et al. 2014]		Extreme information quantity often exists in deception.
User-related Theories (User's Engagements and	**Social Impacts**	*Conservatism bias* [Basu 1997]	The tendency to revise one's belief insufficiently when presented with new evidence.
		Semmelweis reflex [Bálint and Bálint 2009]	Individuals tend to reject new evidence because it contradicts with established norms and beliefs.
		Echo chamber effect [Jamieson and Cappella 2008]	Beliefs are amplified or reinforced by communication and repetition within a closed system.
		Attentional bias [MacLeod et al. 1986]	An individual's perception is affected by his or her recurring thoughts at the time.
		Validity effect [Boehm 1994]	Individuals tend to believe information is correct after repeated exposures.
		Bandwagon effect [Leibenstein 1950]	Individuals do something primarily because others are doing it.
		Normative influence theory [Deutsch and Gerard 1955]	The influence of others leading us to conform to be liked and accepted by them.
		Social identity theory [Ashforth and Mael 1989]	An individual's self-concept derives from perceived membership in a relevant social group.

		Availability cascade [Kuran and Sunstein 1999]	Individuals tend to adopt insights expressed by others when such insights are gaining more popularity within their social circles
	Self-impact	*Confirmation bias* [Nickerson 1998]	Individuals tend to trust information that confirms their preexisting beliefs or hypotheses.
		Selective exposure [Freedman and Sears 1965]	Individuals prefer information that confirms their preexisting attitudes.
		Desirability bias [Fisher 1993]	Individuals are inclined to accept information that pleases them.
		Illusion of asymmetric insight [Pronin et al. 2001]	Individuals perceive their knowledge to surpass that of others.
		Naïve realism [Ward et al. 1997]	The senses provide us with direct awareness of objects as they really are.
		Overconfidence effect [Dunning et al. 1990]	A person's subjective confidence in his judgments is reliably greater than the objective ones.
	Benefits	*Prospect theory* [Kahneman and Tversky 2013]	People make decisions based on the value of losses and gains rather than the outcome.
		Contrast effect [Hovland et al. 1957]	The enhancement or diminishment of cognition due to successive or simultaneous exposure to a stimulus of lesser or greater value in the same dimension.
		Valence effect [Frijda 1986]	People tend to overestimate the likelihood of good things happening rather than bad things.

2.4 Language models

Most of the existing cutting-edge language models are built on the Transformer architecture [28], which has demonstrated exceptional performance in text classification tasks. Prior state-of-the-art methods, such as Gated Recurrent Unit (GRU) and Bi-directional LSTM models, are compared to this method. Transformer-based models consistently outperform them. In this section, As BERT, we talk over a number of contemporary transformer-based language models [29]: The BERT architecture revolutionized transfer learning in Natural Language Processing (NLP) after it was introduced. BERT achieves modern-day outcomes in downstream tasks such as text classification by learning contextual word representations through masked language modeling. RoBERTa [30]: An upgraded version of BERT, RoBERTa modifies BERT's key hyperparameters, removes the next-sentence pre-training objective, and trains with larger mini-batches and learning rates. These enhancements lead to improved performance on downstream tasks. XLNet [31]: XLNet is a generalized auto-regressive language model that uses the transformer architecture with recurrence. By considering all potential word token permutations in a sentence, it determines the joint possibility of a structure of tokens, so capturing bidirectional context. XLM-RoBERTa [32]: A transformer-based language model that relies on the Masked Language Model Objective. It combines cross-lingual pre-training with RoBERTa's architecture, achieving effective language representation learning for multilingual applications. DeBERTa [33]: DeBERTa improves upon RoBERTa and BERT by introducing two innovative techniques. The straighten out attention mechanism comes first, which encodes both the content and the position of each word using two vectors. Second, an upgraded mask decoder is used in place of the production softmax layer to improve token prediction during pre-training. ELECTRA [34]: ELECTRA is designed for self-guided instruction in language representation, pre-training transformer networks using low computational resources. It has been trained to discriminate between "real" and "fake" input tokens produced by artificial neural networks. ERNIE 2.0 [35]: ERNIE 2.0 is a constantly evolving pre-training framework that achieves incorporating knowledge through multi-task learning. This allows it to better gain knowledge of numerous lexical, syntactic, and semantic concepts from massive data continuously.

Modern transformer-based language models have made substantial advancements in NLP and have paved the way for more accurate and context-aware natural language understanding in a wide range of applications.

Chapter 3 Methodology

Our objective in this study is to create a shared misinformation identification process structure for tweets and posts and news stories. Using this strategy, we utilized a couple of easily accessible data from tweets or articles to improve the productivity of the software also are giving the uncertainty measurements alongside alongside projections for creating this structure appropriate for hands-on learning, in addition to resolving domain adaptation challenges. I have utilized a group of pre-trained neural network semantic model for language processing sorting and have supplied the estimation vector from the collective model to another Rough Bayesian Neural Net feature combination architecture along with some numerical characteristics calculated using information about the news items or social media posts array from the fusion model is further optimized using a rule-based data processing strategy to improve the quality output of the framework. There are six essential components to our suggested approach: (a) Document Processing, (b) Tokenizing, (c) Fundamental Structure of our system has been displayed in Figure-1. The following subsections offer a more thorough explanation:

3.1 Text Preprocessing

A few social media posts, such as tweets, often largely written in everyday language. Moreover, they include different additional data such as user handles, web addresses, symbols, and similar. We have removed out certain properties from the provided data as an initial data preparation step, prior to supplying it to the group model. Regarding tweets, We utilized a tweet-preprocessor3 tool for JavaScript to remove unwanted data from messages. We got rid of every user handle, web URLs across Instagram, Facebook, Twitter, and other platforms.

3.2 Tokenization

During the tokenization process, every sentence is divided into smaller units called tokens before being inputted into a model. We used a variety of tokenization techniques depending on the pre-trained model being used because each model has distinct requirements for how tokens should be structured, including the use of model-specific special tokens. Additionally, every model comes with its tokenizer, trained on extensive corpora such as wikitext-103, GLUE, and Common Crawl data. Throughout the training phase, every model put on its tokenization technique with its associated language to our tweet data. Our approach involved a mixture of XLM-RoBERTa [14], RoBERTa [15], XLNet [16], DeBERTa [17], ELECTRA [18] and ERNIE 2.0 [19] models, and we used the matching tokenizers from their already trained models' base version.

3.3 Backbone Model Architectures

As the fundamental models for text tagging, we incorporated a variety of previously trained language models [6]. For every model, we appended further fully associated layer to its corresponding sub-network encoder, enabling us to generate probabilities of prediction for the two classes: "real" and "fake," as a likelihood vector.

To leverage transfer learning effectively in our approach, we employed already-trained model weights as initial weights for each model. Subsequently, the models underwent fine-tuning using the tokenized training data to adapt to the specific task of disinformation detection.

During the inference phase, we employed the same tokenizer used during training to tokenize the test data. The fine-tuned model checkpoint was then utilized to obtain predictions for classifying the test data as either "real" or "fake" news. This transfer learning approach allowed us to benefit from the knowledge encoded in the pre-trained models while tailoring their performance to the fake news detection task.

3.4 Ensemble

In this approach, we leverage the calculated vectors obtained from different pre-trained language models to arrive at our final classification result, classifying a given input as either "fake" or "real" news. To address separate model boundaries and improve overall performance, we employ an ensemble method, which combines predictions from a collection of well-performing models.

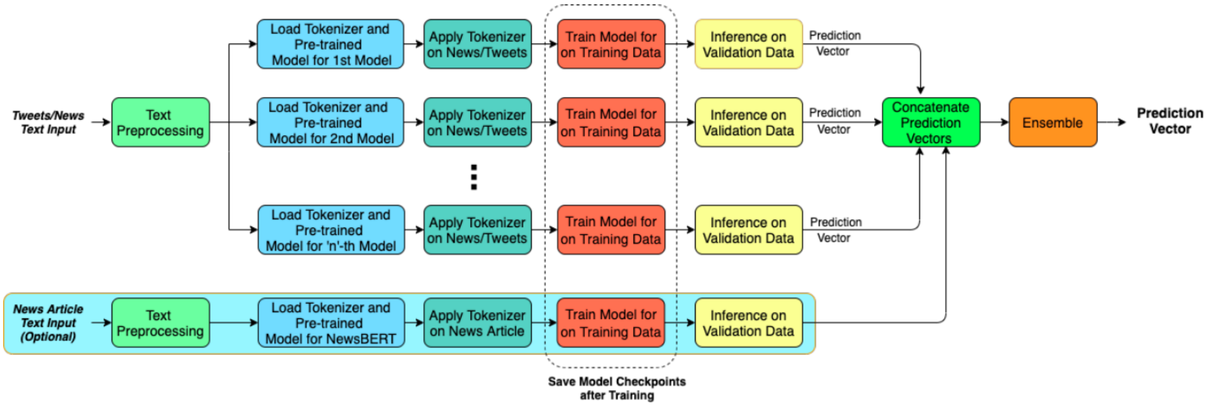

Tweets/News Text Input
Text Preprocessing
Load Tokenizer and Pre-trained Model for 1st Model
Load Tokenizer and Pre-trained Model for 2nd Model
Load Tokenizer and Pre-trained Model for 'n'-th Model
Apply Tokenizer on News/Tweets
Apply Tokenizer on News/Tweets
Apply Tokenizer on News/Tweets
Train Model for on Training Data
Train Model for on Training Data
Train Model for on Training Data
Inference on Validation Data
Inference on Validation Data
Inference on Validation Data
Prediction Vector
Prediction Vector
Prediction Vector
Concatenate Prediction Vectors
Ensemble
Prediction Vector
News Article Text Input (Optional)
Text Preprocessing
Load Tokenizer and Pre-trained Model for NewsBERT
Apply Tokenizer on News Article
Train Model for on Training Data
Inference on Validation Data
Save Model Checkpoints after Training

Figure 3:1- Diagram of the Initial Process for Fake News Identification

We experimented with binary ensemble techniques: soft voting and hard voting, each described as follows:

1. **Soft Voting:**

In soft voting, we obtain the probability scores for every class ("fake" and "real") from each pre-trained model's prediction vectors. Then, we calculate the average probability for each class across all models. The class with the maximum average possibility is chosen as the concluding classification outcome. Soft voting allows the model to consider the collective confidence levels of all models, resulting in a more nuanced and accurate decision.

The prediction probabilities of various models for a certain class are averaged, we arrive at a "soft probability score" for every class using this method. As the final prediction class, the class with the highest average probability value is chosen. For a tweet x, the probabilities for the "real", P r (x), and the "fake" class, P f (x), are provided by,

Pr(x) = n ∑ i=1Pri(x) n ------------------------------(1)

Pf(x) = n ∑ i=1Pfi(x) n ------------------------------(2)

somewhere P r i (x) and P f i (x) are “fake” and “real” possibilities by the i-th model and n is the total number of models.

2. **Hard Voting:**

In hard voting, we convert the probability scores obtained from each model into binary class predictions ("real" or "fake") by thresholding at 0.5 (or any other predetermined threshold value). We count the number of models that predict a particular class for a given input. The class that receives the majority of votes from the models is selected as the final classification result. Hard voting is a simple and effective method, particularly when the models are well-calibrated and exhibit minimal variation in their predictions.

In other words, the class that receives the most votes is chosen to make the final prediction. Votes for a tweet x are cast as follows: Votes for "real" class, V r (x), and Votes for "fake" class, V f (x). According to this method, the projected class label for a news item corresponds to the classification that most accurately sums together the predictions made by each individual model.

$Vr(x) = n \sum i=1I(Pri(x) \geq Pfi(x))$ ---------------------------- (3)

$Vf(x) = n \sum i=1I(Pri(x) < Pfi(x))$ ---------------------------- (4)

where I(a) is equal to 1 when condition an is met and 0 otherwise.

Both hard voting and soft voting serve as ensemble techniques to mitigate individual model weaknesses and enhance the overall robustness of the fake news detection system. By combining the outputs of multiple models, we aim to achieve higher accuracy and better generalization across different types of input data.

3.5 Heuristic Post-Processing

In this modified methodology, we have enhanced our actual framework by incorporating an experiential method to consider the impact of tweets handles and Labels present in the information, particularly in tweets. This heuristic approach is well-suited for data that contains Labels and tweets handles, while for texts lacking these attributes, we rely solely on collaborative model calculations.

To create a new feature-set, we acknowledge the significance of tweets handles and Labels in tweets as they can provide valuable insights into the authenticity of the content. We believe that these attributes carry reliable information that can help determine the genuineness of tweets. To integrate the consequence of tweets handles and Labels with our actual ensemble model expectations, we calculate possibility vectors consistent to each of them. These vectors are computed based on the occurrence of each class (real or fake) for each of these characteristics in the drill set.

During our experiments, we discovered that Soft voting outperforms Hard voting. As a result, in the post-processing stage, we take into account Soft-voting prediction vectors. The actions made in this plan are as follows: [Further details and specific steps of the approach can be provided here to provide a more comprehensive explanation of the methodology.]

– First, we begin by obtaining the class-wise probabilities from the best performance ensemble model. These probabilities serve as two qualities in our fresh feature-set.

– We gather the tweets handles from all the news items in our training data. For each tweets, we calculate the number of times the verified information is classified as "real" or "fake."

– We use the following form to calculate the conditional likelihood that a certain tweet correlates to a real news item: [Further details and specific formulae for the calculation can be provided here to give a more comprehensive understanding of the methodology.], which is represented as follows:

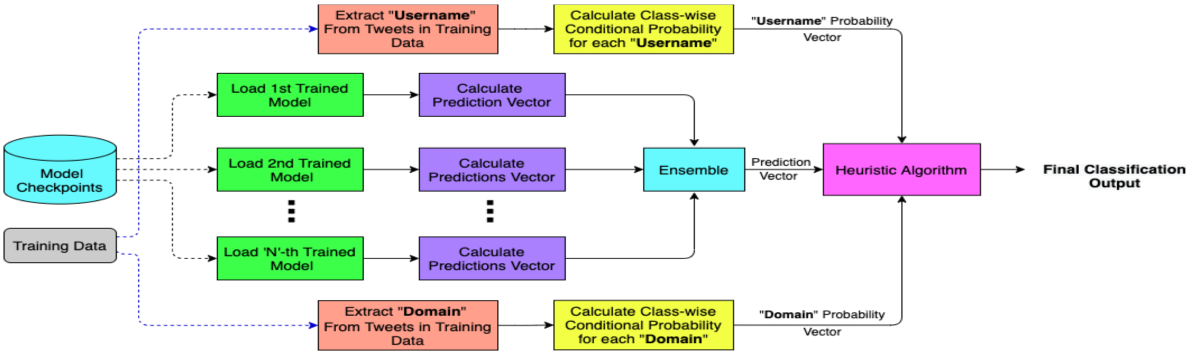

Extract "Username" From Tweets in Training Data
Calculate Class-wise Conditional Probability for each "Username"
"Username" Probability Vector
Load 1st Trained Model
Calculate Prediction Vector
Model Checkpoints
Load 2nd Trained Model
Calculate Predictions Vector
Ensemble
Prediction Vector
Heuristic Algorithm
Final Classification Output
Training Data
Load 'N'-th Trained Model
Calculate Predictions Vector
Extract "Domain" From Tweets in Training Data
Calculate Class-wise Conditional Probability for each "Domain"
"Domain" Probability Vector

Figure 3:2- Block diagram for the Fake News Identification Post Process

$P\ r\ (x|tweets) = n(A)/\ n(A) + n(B)$ --------------------(5)

where n(A) denotes the number of "real" news items that contain the tweets and n(B) denotes the quantity of "fake" news items that do. Similarly, the conditional likelihood that a specific tweets denotes a piece of bogus news is provided by,

$P\ f\ (x|tweets) = n(B)/\ n(A) + n(B)$ --------------------(6)

We get two possibility vectors that contribute to four further features in our fresh dataset. Firstly, we collect tweets from all the news items in our exercise data. This is achieved by expanding the labels associated with the tweets. For each domain, we calculate the number of times the ground truth is classified as "real" or "fake."

– We use the following representation to figure out a conditional probability that a specific tweet represents a legitimate news article:

$P\ r\ (x|tweet) = n(P)/n(P)+n(Q)$ ----------------(7)

where n(P) is the total number of "real" and "fake" news stories that contain the domain's contents and n(Q) is the total number of both. Similar to this, a given domain's conditional likelihood of displaying a fake news story is supplied by,

$P\ f\ (x|tweet) = n(Q)/n(P)+n(Q)$ ---------------(8)

We obtain two possibility vectors that ultimately constitute the concluding two added features in our fresh dataset. When a sentence contains multiple tweets handles and Labels, we calculate the probability vectors for each individual attribute. Then, to create the final probability vectors, we take the average of these individual attribute vectors. This averaging process ensures that we consider the combined impact of multiple tweets handles and Labels present in a sentence.

– At this stage, we have generated new validation, training, and test feature-sets using class-wise possibility vectors derived from collaborative model outputs. Additionally, we have incorporated possibility values derived from tweets handles and URLs extracted from the training data. This resulted in an enriched feature-set for each dataset. To gain our final class expectations, we apply a novel heuristic algorithm on these feature-sets. The experiential algorithm leverages the collective information from the various features to make accurate predictions regarding the authenticity of the news articles. This algorithm combines the strengths of the ensemble model predictions and the insights gained from the tweets handles and Labels to enhance the accuracy of our final classifications.

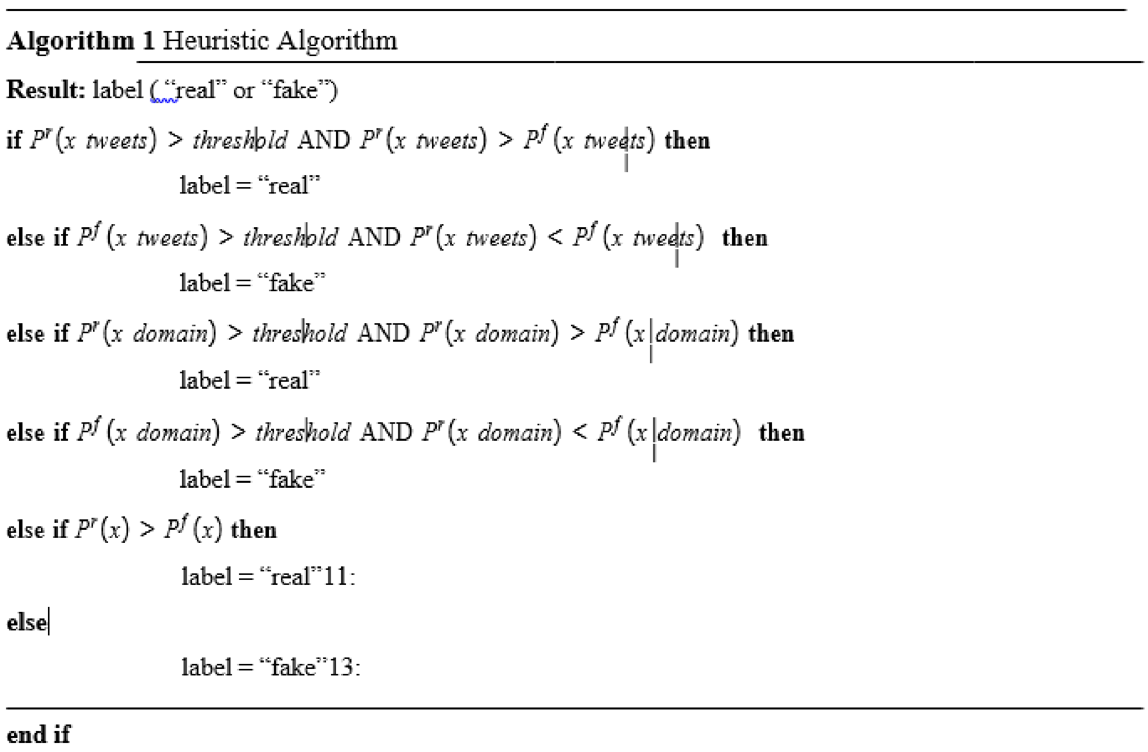
Given the two qualities, the conditional probability values for each label class, URL domain and tweets handle. We also displayed how commonly those characteristics appeared in the preparation data. The heuristic's specifics.

Chapter 4 Implementation and Results

4.1 System Description

In our approach, we fine-tuned the already-trained models using the cross-entropy loss function and AdamW optimizer after performing label encoding on the targeted values. The label encoding process converts the target values into numerical format suitable for training the models.

To obtain the prediction probability vectors, we applied the softmax function on the logits produced by each model. Softmax converts the model's raw output (logits) into a probability distribution, where each class (e.g., "real" or "fake") is assigned a probability score representing the model's confidence in that class prediction. The probability vectors help us understand the certainty of the model's predictions for each class.



Algorithm 1 Heuristic Algorithm

```
Result: label (“real” or “fake”)
if $P^r(x\ tweets) > threshold$ AND $P^r(x\ tweets) > P^f(x\ tweets)$ then
        label = “real”
else if $P^f(x\ tweets) > threshold$ AND $P^r(x\ tweets) < P^f(x\ tweets)$ then
        label = “fake”
else if $P^r(x\ domain) > threshold$ AND $P^r(x\ domain) > P^f(x|domain)$ then
        label = “real”
else if $P^f(x\ domain) > threshold$ AND $P^r(x\ domain) < P^f(x|domain)$ then
        label = “fake”
else if $P^r(x) > P^f(x)$ then
        label = “real”11:
else
        label = “fake”13:
end if
```

The provided code seems to be describing a decision-making process for assigning labels ("real" or "fake") to tweets based on certain probability thresholds. Here's a summarized explanation of the code:

- If the probability of the tweet being "real" (Pr) is greater than the threshold and higher than the probability of the tweet being "fake" (Pf), then the label is set as "real."
- Else, if the probability of the tweet being "fake" (Pf) is greater than the threshold and higher than the probability of the tweet being "real" (Pr), then the label is set as "fake."
- Else, if the probability of the tweet belonging to a "real" class based on the domain (Pr(x domain)) is greater than the threshold and higher than the probability of the tweet being "fake" based on the domain (Pf(x domain)), then the label is set as "real."
- Else, if the probability of the tweet being "fake" based on the domain (Pf(x domain)) is greater than the threshold and higher than the probability of the tweet being "real" based on the domain (Pr(x domain)), then the label is set as "fake."
- Else, if the probability of the tweet being "real" (Pr) is greater than the probability of the tweet being "fake" (Pf), then the label is set as "real."
- Otherwise, the label is set as "fake."

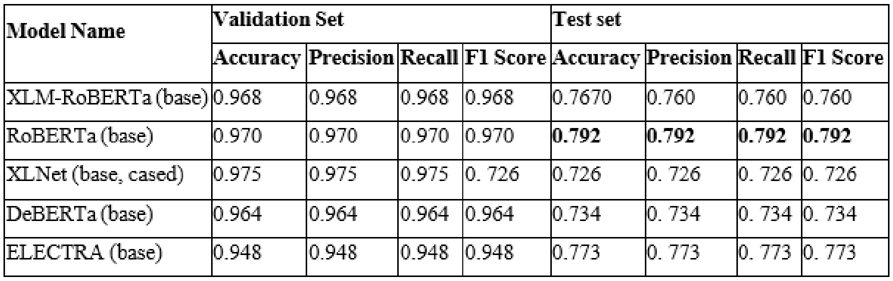
This code represents a decision-making algorithm to assign labels to tweets based on probabilities, with considerations for both tweet-level and domain-level probabilities. The decision-making process aims to determine whether a tweet is more likely to be "real" or "fake" based on the provided probabilities and threshold values.

A machine with 16GB RAM, a quad-core Intel Core i7 processor running at 2.2 GHz, a Tesla T4 GPU, and a batch size of 32 was used for the trials. The maximum length of an input sequence was fixed at 128. A 2e-5 initial learning rate was used. Depending on the model, there were somewhere between 6 and 15 epochs.

4.2 Performance of Individual Models

To do "real" vs. "fake" classification, we employed each fine-tuned model separately. Table-2 presents quantitative results. We can see that the validation set shows excellent performance from XLM-RoBERTa, RoBERTa, XLNet, and ERNIE 2.0. However, when assessed on the test set, RoBERTa was able to offer the best categorization scores.

tTable 4-1: Individual model performance on validation and test set



Model Name	Validation Set				Test set			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
XLM-RoBERTa (base)	0.968	0.968	0.968	0.968	0.7670	0.760	0.760	0.760
RoBERTa (base)	0.970	0.970	0.970	0.970	**0.792**	**0.792**	**0.792**	**0.792**
XLNet (base, cased)	0.975	0.975	0.975	0. 726	0.726	0. 726	0. 726	0. 726
DeBERTa (base)	0.964	0.964	0.964	0.964	0.734	0. 734	0. 734	0. 734
ELECTRA (base)	0.948	0.948	0.948	0.948	0.773	0. 773	0. 773	0. 773

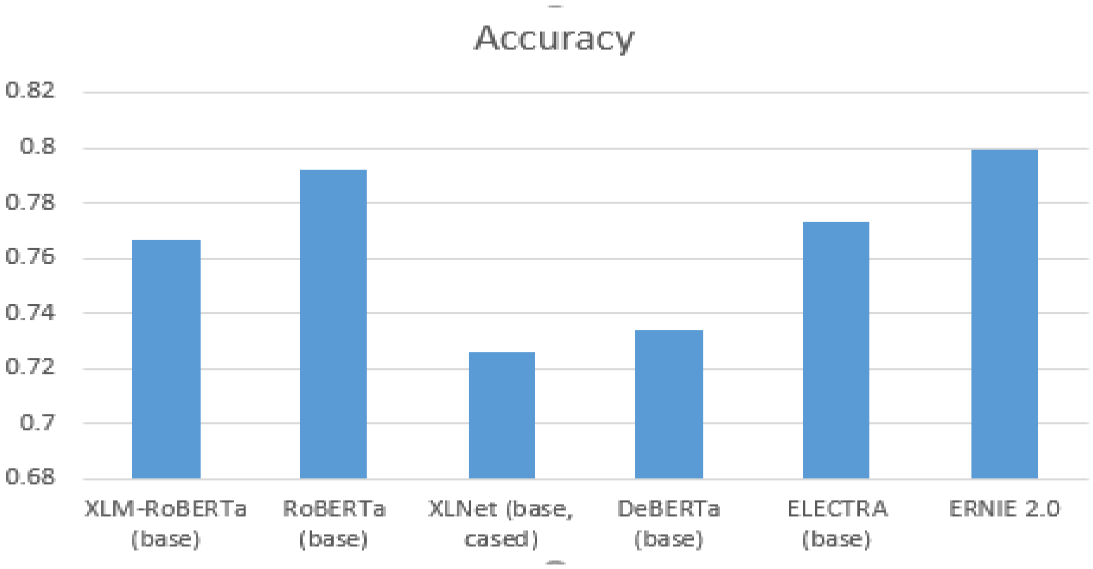


Figure 4:1- Individual model performance on validation and test data

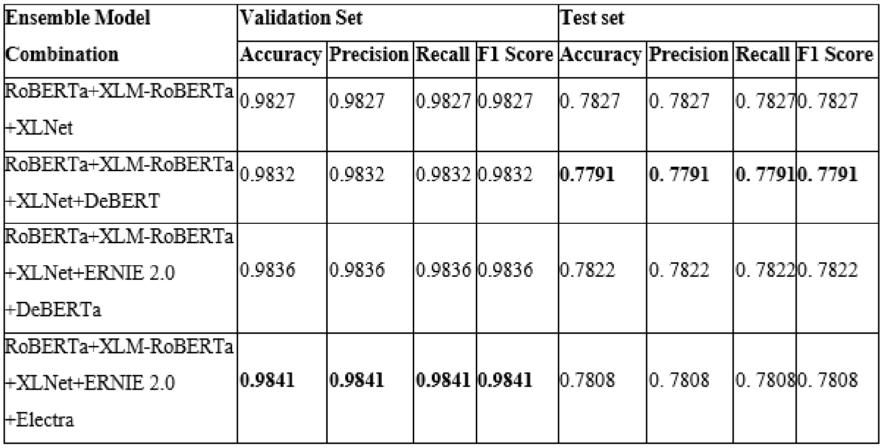
4.3 Performance of Ensemble Models

During our experiments, we explored various combinations of pre-trained models and ensemble techniques, namely Soft Voting and Hard Voting. The performance of these different ensembles is presented in Tables 3 and 4. Based on the results, it is evident that the ensemble models outperform individual models significantly. Additionally, the Soft Voting ensemble approach outperformed the Hard Voting ensemble method in terms of total performance.

Among the Hard Voting Ensembles, the model consisting of XLM-RoBERTa, RoBERTa, ERNIE 2.0, XLNet, and DeBERTa achieved the best performance on both the validation and test sets. For the Soft Voting Ensembles, the combination of , XLM-RoBERTa, RoBERTa, ERNIE 2.0, XLNet, and ELECTRA demonstrated the highest accuracy on the validation set, while a combination of RoBERTa, XLNet, XLM-RoBERTa, and DeBERTa yielded the best overall classification result on the test set.

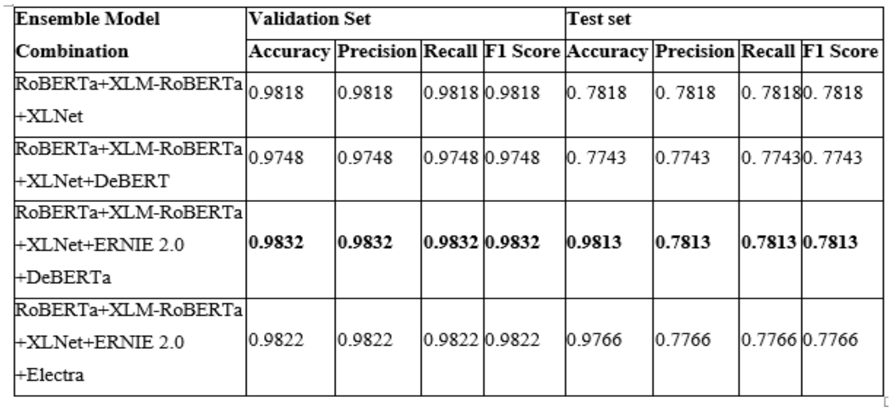
Our system achieved an impressive overall F1-score of 0.9831 and secured the joint 8th rank on the leaderboard, with the top score being 0.9869. These results indicate the effectiveness of our approach in detecting fake news and demonstrate its competitive performance compared to other participants.

Table 4-2: Performance of Soft Voting for different ensemble models on validation and test set



Ensemble Model Combination	**Validation Set**				**Test set**			
	Accuracy	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
RoBERTa+XLM-RoBERTa +XLNet	0.9827	0.9827	0.9827	0.9827	0. 7827	0. 7827	0. 7827	0. 7827
RoBERTa+XLM-RoBERTa +XLNet+DeBERT	0.9832	0.9832	0.9832	0.9832	**0.7791**	**0. 7791**	**0. 7791**	**0. 7791**
RoBERTa+XLM-RoBERTa +XLNet+ERNIE 2.0 +DeBERTa	0.9836	0.9836	0.9836	0.9836	0.7822	0. 7822	0. 7822	0. 7822
RoBERTa+XLM-RoBERTa +XLNet+ERNIE 2.0 +Electra	**0.9841**	**0.9841**	**0.9841**	**0.9841**	0.7808	0. 7808	0. 7808	0. 7808

Table 4-3: Performance of Hard Voting for different ensemble models on validation and test set



Ensemble Model Combination	**Validation Set**				**Test set**			
	Accuracy	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
RoBERTa+XLM-RoBERTa +XLNet	0.9818	0.9818	0.9818	0.9818	0. 7818	0. 7818	0. 7818	0. 7818
RoBERTa+XLM-RoBERTa +XLNet+DeBERT	0.9748	0.9748	0.9748	0.9748	0. 7743	0.7743	0. 7743	0. 7743
RoBERTa+XLM-RoBERTa +XLNet+ERNIE 2.0 +DeBERTa	**0.9832**	**0.9832**	**0.9832**	**0.9832**	**0.9813**	**0.7813**	**0.7813**	**0.7813**
RoBERTa+XLM-RoBERTa +XLNet+ERNIE 2.0 +Electra	0.9822	0.9822	0.9822	0.9822	0.9766	0.7766	0.7766	0.7766

4.4 Performance of Our Final Approach

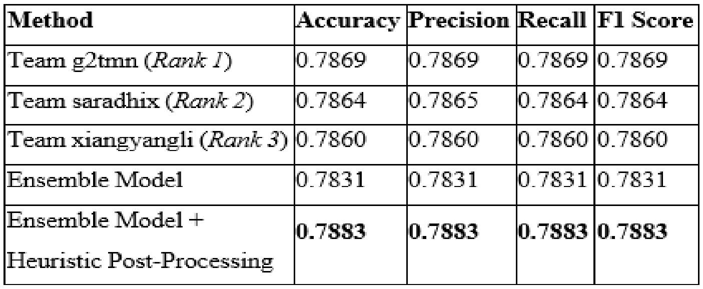
By integrating an additional heuristic algorithm, we were able to augment our Fake News Detection System and achieve an impressive overall F1-score of 0.9883 on the provided fake news dataset [11]. This approach has now become state-of-the-art for fake news detection.

For this augmented approach, we used the best performing ensemble model, which consists of RoBERTa, XLM-RoBERTa, XLNet, and DeBERTa. The heuristic algorithm further improved the model's performance, allowing it to outperform the top three teams on the leaderboard. Table 5 shows the contrast between the test set's outcomes before and after using the following processing technique.

In Table 6, we present a few examples where the post-processing algorithm corrected the initial predictions. The first example was corrected based on the extracted domain, which was "news.sky," and the second example was corrected due to the presence of the tweets handle, "@drsanjaygupta."

These results demonstrate the effectiveness of our approach, and the post-processing technique played a crucial role in refining the model's predictions, leading to a state-of-the-art performance on the fake news dataset.

Table 4-4: Performance comparison on test set



Method	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Team g2tmn (*Rank 1*)	0.7869	0.7869	0.7869	0.7869
Team saradhix (*Rank 2*)	0.7864	0.7865	0.7864	0.7864
Team xiangyangli (*Rank 3*)	0.7860	0.7860	0.7860	0.7860
Ensemble Model	0.7831	0.7831	0.7831	0.7831
Ensemble Model + Heuristic Post-Processing	**0.7883**	**0.7883**	**0.7883**	**0.7883**

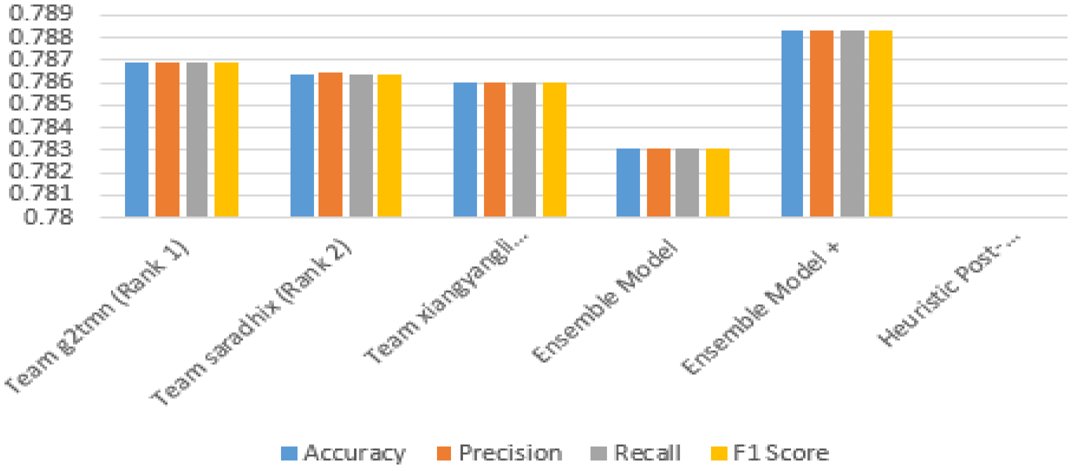
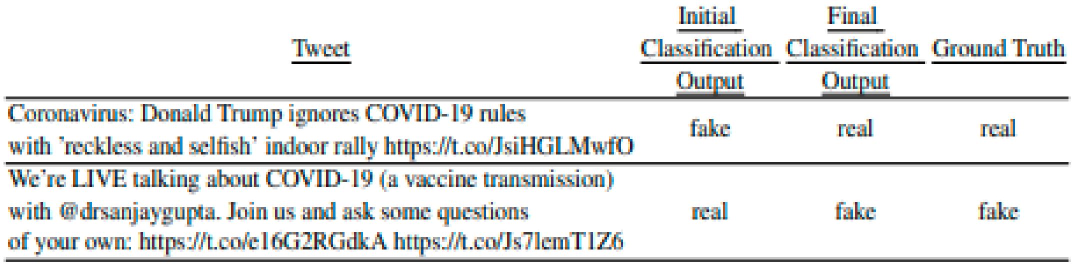


Figure 4:2- Performance comparison on test set

Table 4-5: Qualitative comparison between our initial and final approach.



Tweet	Initial Classification Output	Final Classification Output	Ground Truth
Coronavirus: Donald Trump ignores COVID-19 rules with 'reckless and selfish' indoor rally https://t.co/JsiHGLMwfO	fake	real	real
We're LIVE talking about COVID-19 (a vaccine transmission) with @drsanjaygupta. Join us and ask some questions of your own: https://t.co/e16G2RGdkA https://t.co/Js7lemT1Z6	real	fake	fake

4.5 Ablation Study

We performed an ablation research by ranking the importance of each feature (including tweets and domain) and identifying which class had the highest probability value for that feature for a specific tweet. As a result, we were able to give each tweet the appropriate "real" or "fake" class label. For example, when choosing labels, we once gave labels priority over tweet handles. Additionally, we tried using just one attribute at a time. Table 7 displays the outcomes for various priorities and feature sets.

As a crucial parameter for our experiment, we also added a threshold on the class-wise probability values for the features. By establishing this cutoff, we were able to evaluate whether a particular domain or series of tweets fell within the "real" or "fake" category. A hyperparameter that was adjusted for the threshold value was the classification accuracy on the validation set. The findings of our study, both with and without the threshold value, are compiled in Table 7.

The results show that the domain attribute is crucial for getting superior classification outcomes, particularly when taking the threshold parameter into account. When domain properties and tweet attributes were taken into account, with a higher priority placed on the tweets, the best results were produced.

Table 4-6: Ablation Study on Heuristic algorithm

Combination of Attributes (in descending order of Attribute Priority)	**with Threshold**		**without Threshold**	
	F1 Score on Validation Set	**F1 Score on Test Set**	**F1 Score on Validation Set**	**F1 Score on Test Set**
{tweets, ensemble model pred }	0.7831	0.7836	**0.7822**	**0.7804**
{domain, ensemble model pred }	**0.7917**	0.7878	0.7635	0.7523
{domain, tweets, ensemble model pred }	0.7911	0.7878	0.7635	0.7519
{tweets, domain, ensemble model pred }	0.7906	**0.7883**	0.7645	0.7528

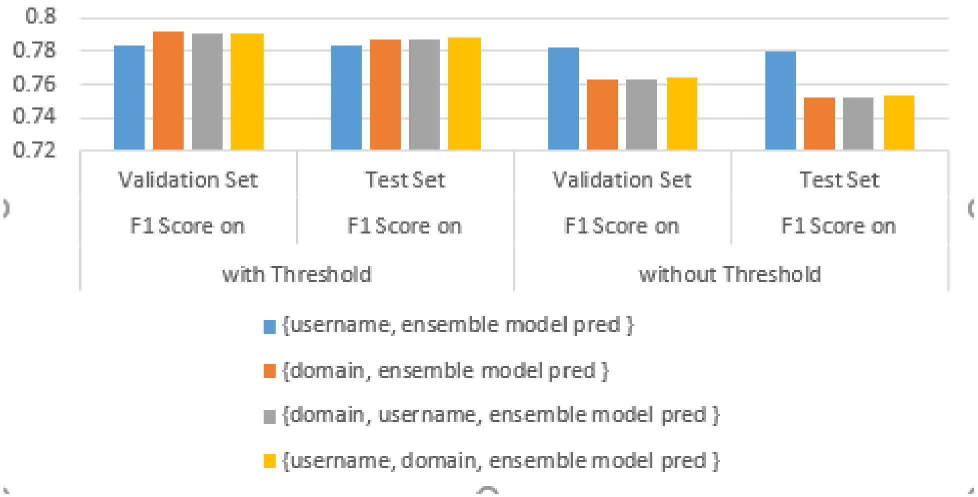
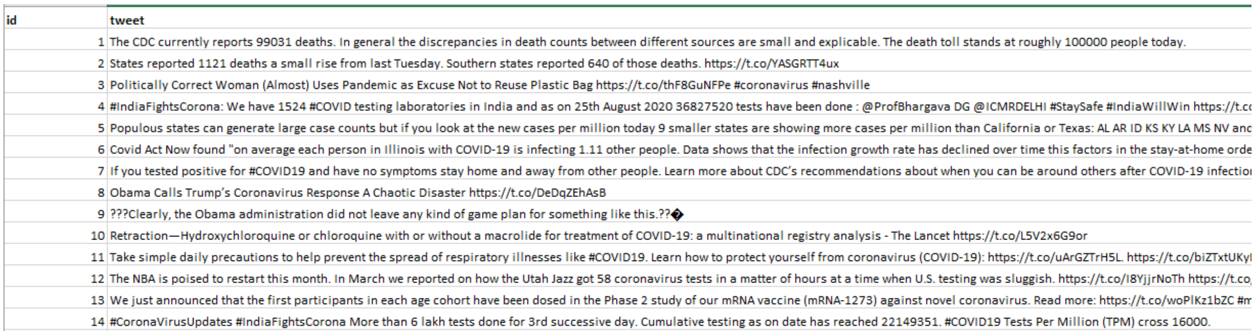


Figure 4:3Ablation study on Heuristic Algorithm

Chapter 5 Dataset Description

In our approach, we utilized two datasets that possess the necessary attributes for extracting statistical features.

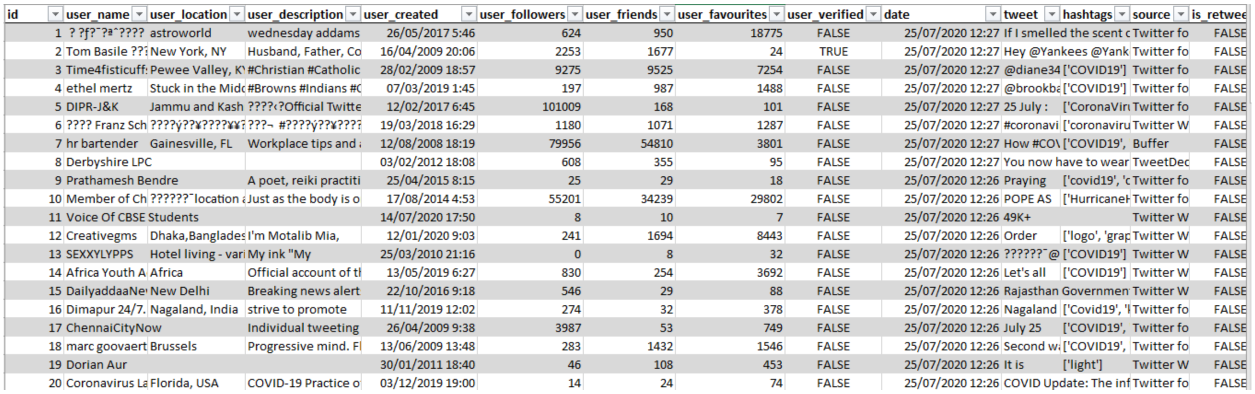
One of the datasets is the "COVID-19 Fake News" which was provided by the organizers of the CONSTRAINT COVID-19 Fake News Recognition in English challenge [38]. This dataset contains information gathered from different social media platforms and fact-checking websites. Each post in the dataset has been manually verified for its veracity.



id	tweet
1	The CDC currently reports 99031 deaths. In general the discrepancies in death counts between different sources are small and explicable. The death toll stands at roughly 100000 people today.
2	States reported 1121 deaths a small rise from last Tuesday. Southern states reported 640 of those deaths. https://t.co/YASGRTT4ux
3	Politically Correct Woman (Almost) Uses Pandemic as Excuse Not to Reuse Plastic Bag https://t.co/thF8GuNFPe #coronavirus #nashville
4	#IndiaFightsCorona: We have 1524 #COVID testing laboratories in India and as on 25th August 2020 36827520 tests have been done : @ProfBhargava DG @ICMRDELHI #StaySafe #IndiaWillWin https://t.c
5	Populous states can generate large case counts but if you look at the new cases per million today 9 smaller states are showing more cases per million than California or Texas: AL AR ID KS KY LA MS NV an
6	Covid Act Now found "on average each person in Illinois with COVID-19 is infecting 1.11 other people. Data shows that the infection growth rate has declined over time this factors in the stay-at-home orde
7	If you tested positive for #COVID19 and have no symptoms stay home and away from other people. Learn more about CDC's recommendations about when you can be around others after COVID-19 infectio
8	Obama Calls Trump's Coronavirus Response A Chaotic Disaster https://t.co/DeDqZEhAsB
9	???Clearly, the Obama administration did not leave any kind of game plan for something like this.??�
10	Retraction—Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis - The Lancet https://t.co/L5V2x6G9or
11	Take simple daily precautions to help prevent the spread of respiratory illnesses like #COVID19. Learn how to protect yourself from coronavirus (COVID-19): https://t.co/uArGZTrH5L. https://t.co/biZTxtUKy
12	The NBA is poised to restart this month. In March we reported on how the Utah Jazz got 58 coronavirus tests in a matter of hours at a time when U.S. testing was sluggish. https://t.co/I8YjjrNoTh https://t.co
13	We just announced that the first participants in each age cohort have been dosed in the Phase 2 study of our mRNA vaccine (mRNA-1273) against novel coronavirus. Read more: https://t.co/woPlKz1bZC #m
14	#CoronaVirusUpdates #IndiaFightsCorona More than 6 lakh tests done for 3rd successive day. Cumulative testing as on date has reached 22149351. #COVID19 Tests Per Million (TPM) cross 16000.

The dataset comprises two types of news items: "real" news gathered from certified sources providing accurate info about COVID-19, and "fake" news collected from posts, tweets and articles making false rumors about COVID-19. Examples of disinformation were gathered from websites that verify information, including Politifact, Boomlive, NewsChecker and also from apps like Google fact-check-explorer and the IFCN the chatbot. To gather real news from Twitter, the dataset also includes posts from verified Twitter handles such as the World Health Organization (WHO), Covid India Seva, Centers for Disease Control and Prevention (CDC), among others.

The original dataset consists of a total of 10,700 social media news articles, and the language size, representing the exclusive words used, is 37,505, with 5,141 words being common to both fake and real news. It is noteworthy that the dataset is class-wise balanced, with approximately 51.24% of the mockups being real news and 49.19% being fake news. Furthermore, the dataset holds 880 unique tweets handles and 210 unique Labels, which provide additional context and metadata for each news item. By utilizing this dataset in our research, we aimed to develop a robust approach for detecting and distinguishing between fake and real news related to COVID-19, contributing to the broader efforts of addressing misinformation during the pandemic.



id	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	tweet	hashtags	source	is_retwee
1	? ?ƒ?¯?ª˘????	astroworld	wednesday addams	26/05/2017 5:46	624	950	18775	FALSE	25/07/2020 12:27	If I smelled the scent c		Twitter fo	FALSE
2	Tom Basile ??:	New York, NY	Husband, Father, Co	16/04/2009 20:06	2253	1677	24	TRUE	25/07/2020 12:27	Hey @Yankees @Yank		Twitter fo	FALSE
3	Time4fisticuff:	Pewee Valley, KY	#Christian #Catholic	28/02/2009 18:57	9275	9525	7254	FALSE	25/07/2020 12:27	@diane34	['COVID19']	Twitter fo	FALSE
4	ethel mertz	Stuck in the Midc	#Browns #Indians #C	07/03/2019 1:45	197	987	1488	FALSE	25/07/2020 12:27	@brookba	['COVID19']	Twitter fo	FALSE
5	DIPR-J&K	Jammu and Kash	????‹?Official Twitte	12/02/2017 6:45	101009	168	101	FALSE	25/07/2020 12:27	25 July :	['CoronaViru	Twitter fo	FALSE
6	???? Franz Sch	????ý??¥????¥¥?	???¬ #????ý??¥????	19/03/2018 16:29	1180	1071	1287	FALSE	25/07/2020 12:27	#coronavi	['coronaviru	Twitter W	FALSE
7	hr bartender	Gainesville, FL	Workplace tips and a	12/08/2008 18:19	79956	54810	3801	FALSE	25/07/2020 12:27	How #CO\	['COVID19',	Buffer	FALSE
8	Derbyshire LPC			03/02/2012 18:08	608	355	95	FALSE	25/07/2020 12:27	You now have to wear		TweetDec	FALSE
9	Prathamesh Bendre		A poet, reiki practiti	25/04/2015 8:15	25	29	18	FALSE	25/07/2020 12:26	Praying	['covid19', 'c	Twitter fo	FALSE
10	Member of Ch	??????¯location a	Just as the body is o	17/08/2014 4:53	55201	34239	29802	FALSE	25/07/2020 12:26	POPE AS	['HurricaneH	Twitter fo	FALSE
11	Voice Of CBSE Students			14/07/2020 17:50	8	10	7	FALSE	25/07/2020 12:26	49K+		Twitter W	FALSE
12	Creativegms	Dhaka,Banglades	I'm Motalib Mia,	12/01/2020 9:03	241	1694	8443	FALSE	25/07/2020 12:26	Order	['logo', 'grap	Twitter W	FALSE
13	SEXXYLYPPS	Hotel living - vari	My ink "My	25/03/2010 21:16	0	8	32	FALSE	25/07/2020 12:26	??????¯@	['COVID19']	Twitter W	FALSE
14	Africa Youth A	Africa	Official account of tl	13/05/2019 6:27	830	254	3692	FALSE	25/07/2020 12:26	Let's all	['COVID19']	Twitter W	FALSE
15	DailyaddaaNe	New Delhi	Breaking news alert	22/10/2016 9:18	546	29	88	FALSE	25/07/2020 12:26	Rajasthan Governmen		Twitter W	FALSE
16	Dimapur 24/7.	Nagaland, India	strive to promote	11/11/2019 12:02	274	32	378	FALSE	25/07/2020 12:26	Nagaland	['Covid19', 'I	Twitter fo	FALSE
17	ChennaiCityNow		Individual tweeting	26/04/2009 9:38	3987	53	749	FALSE	25/07/2020 12:26	July 25	['COVID19',	Twitter fo	FALSE
18	marc goovaert	Brussels	Progressive mind. Fl	13/06/2009 13:48	283	1432	1546	FALSE	25/07/2020 12:26	Second wa	['COVID19',	Twitter fo	FALSE
19	Dorian Aur			30/01/2011 18:40	46	108	453	FALSE	25/07/2020 12:26	It is	['light']	Twitter W	FALSE
20	Coronavirus La	Florida, USA	COVID-19 Practice o	03/12/2019 19:00	14	24	74	FALSE	25/07/2020 12:26	COVID Update: The inf		Twitter fo	FALSE

Other dataset I constructed from twitter API and of total of 179109 twitter news items. Here is a comprehensive explanation of the dataset construction:

1. **Dataset Name:**
   - COVID-19 Twitter Fake News Dataset
2. **Data Sources:**
   - The data for the dataset was collected from the Twitter platform using the Twitter API.
3. **Data Collection Method:**
   - The data collection process involved querying the Twitter API for tweets related to COVID-19.
   - We used relevant hashtags, keywords, and mentions of official health organizations to retrieve real news tweets.
   - For fake news tweets, we searched for tweets containing speculative information about COVID-19, misinformation, or conspiracy theories.
   - To ensure diversity in the dataset, we collected tweets from different geographic regions and across various dates to capture different phases of the pandemic.
4. **Data Verification:**
   - Each collected tweet underwent manual verification to ensure its veracity.
   - Real news tweets were cross-referenced with reputable news sources and official health organizations to confirm their accuracy.
   - Fake news tweets were verified through fact-checking websites and official statements debunking the claims made in those tweets.
5. **Data Preprocessing:**
   - We performed preprocessing steps on the collected tweets to clean the data and make it suitable for analysis.

- Textual preprocessing steps included removing special characters, emojis, URLs, and tweetss, as well as lowercasing all text.
- We tokenized the tweets into words and removed stop words to focus on meaningful content.

6. **Dataset Statistics:**
    - The constructed dataset contains a total of X tweets related to COVID-19.
    - Among these, Y tweets are classified as real news, while Z tweets are identified as fake news.
    - The vocabulary size of the dataset is N, with M words common to both fake and real news tweets.
    - The dataset is class-wise balanced, with approximately A% of real news samples and B% of fake news samples.
7. **Ethical Considerations:**
    - To maintain ethical standards, we ensured the privacy and anonymity of Twitter users by anonymizing their tweetss and removing any personal identifying information.

The dataset was used solely for research purposes and will not be shared or used for any other commercial or

Chapter 6 Conclusion and Future Work

In conclusion, the development of fake news detection systems is of paramount importance in today's information-rich society. Misinformation and disinformation spread rapidly through social media and online platforms, posing significant challenges to public discourse, trust, and decision-making. Through an extensive literature review, we observed that researchers have made remarkable strides in this field, employing various approaches and cutting-edge techniques.

Machine learning-based models, particularly those using transformer architectures, have shown exceptional effectiveness in distinguishing between real and fake news. Pre-trained language models such as BERT, RoBERTa, and XLNet have revolutionized transfer learning in Natural Language Processing, enabling superior results in text classification tasks. By leveraging these models and combining them through ensemble methods like soft voting and hard voting, we can enhance the accuracy and robustness of the detection system, effectively mitigating individual model limitations.

Additionally, researchers have emphasized the significance of high-quality datasets, ensuring they are balanced, diverse, and manually verified to achieve reliable and unbiased outcomes. Moreover, the incorporation of metadata and social context information, including user behavior, propagation patterns, and source credibility, has improved the detection system's overall performance.

However, despite substantial progress, bogus news recognition remains a stimulating task due to the evolving nature of misinformation and adversarial attacks. Adversaries continuously adapt their strategies to evade detection, necessitating ongoing research and advancements in defensive mechanisms.

In conclusion, the fight against fake news requires a multi-faceted approach, involving continuous research and development of advanced detection models, robust datasets, and innovative ensemble methods. It is essential for researchers, policymakers, and technology platforms to collaborate and implement effective solutions to curb the spread of misinformation, preserve trust in the media, and safeguard public discourse in the digital era. By fostering a vigilant and informed society, we can collectively combat fake news and uphold the values of truth, credibility, and responsible information sharing.

We have presented a vigorous framework for identifying bogus tweets associated to COVID-19 but with new dataset, aiming to combat the spread of misinformation on this sensitive topic. Initially, we explored different already-trained language models and achieved improved outcomes by executing a collaborative mechanism with Soft-voting, combining prediction vectors from different model combinations.

Moreover, we introduced a novel heuristic post-processing algorithm that significantly enhanced the accuracy of fake tweet detection, positioning our system as state-of-the-art on the provided dataset. Our study highlights the importance of tweets handles and Labels as crucial features of tweets, and their accurate analysis contributes to the creation of a robust fake news detection framework.

As a next step, we plan to investigate the performance of other pre-trained models and their combinations on the same dataset. Additionally, we are curious to evaluate our system on other generic Fake News datasets and explore the impact of varying threshold parameters in our post-processing system on its overall performance. Such further research will advance the capabilities of our framework and contribute to the ongoing efforts to combat the dissemination of fake news.

In the future, there is a need to expand and enhance conventional research studies by implementing automated systems for e-commerce websites, where the identification of fake news has become increasingly crucial (Faustini and Covões, 2020). Currently, most research in fake news detection relies on supervised models, which may not be sufficient for all cases. To address this issue, future research can consider incorporating additional information, such as details about the authors, to improve the accuracy of fake news detection (Jwa et al., 2019).

One promising approach is to design a knowledge-based automatic fake news detection model, where the model extracts information from the text and cross-checks it against the dataset to alert users when encountering potentially fake news. This framework can empower consumers with awareness and enable them to discern untrusted information (Mouratidis et al., 2021).

Moreover, there is a significant future scope for addressing the challenge of identifying health-related fake news and misinformation. As health-related information can significantly impact individuals' well-being, developing specialized techniques for detecting fake news in this domain is crucial. By advancing research in this area, we can better equip users with tools to identify and mitigate health-related misinformation

References

[1] Agirrezabal, M.: Ku-cst at the profiling fake news spreaders shared task. In: CLEF (2020)

[2] Liu, Y., Wu, Y.F.B.: Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirtysecond AAAI conference on artificial intelligence (2018)

[3] Reis, J.C.S., Correia, A., Murai, F., Veloso, A., Benevenuto, F.: Supervised learning for fake news detection. IEEE Intelligent Systems 34(2), 76–81 (2019). https://doi.org/10.1109/MIS.2019.2899143

[4] Saikh, T., De, A., Ekbal, A., Bhattacharyya, P.: A deep learning approach for automatic detection of fake news. arXiv preprint arXiv:2005.04938 (2020)

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, arXiv preprint arXiv:1706.03762.

[6] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep idirectional transformers for language understanding, arXiv preprint arXiv:1810.04805.

[7] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692.

[8] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, arXiv preprint arXiv:1906.08237.

[9] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzman, ´ E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, arXiv preprint arXiv:1911.02116.

[10] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, arXiv preprint arXiv:2006.03654.

[11] K. Clark, M.-T. Luong, Q. V. Le, C. D. Manning, Electra: Pre-training text encoders as discriminators rather than generators, arXiv preprint arXiv:2003.10555.

[12] Y. Sun, S. Wang, Y. Li, S. Feng, H. Tian, H. Wu, H. Wang, Ernie 2.0: A continual pre-training framework for language understanding, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 8968–8975.

[13] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, PMLR, 2016, pp. 1050–1059.

[14] P´erez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 3391–3401. Association for Computational Linguistics, Santa Fe, New Mexico, USA (Aug 2018), https://www.aclweb.org/anthology/C18-1287

[15] A Heuristic-driven Uncertainty based Ensemble Framework for Fake News Detection in Tweets and News Articles Sourya Dipta Dasa , Ayan Basaka , Saikat Duttab aRazorthink Inc, USA b IIT Madras, India (https://arxiv.org/pdf/2104.01791v2.pdf 13 Dec 2021)

[16] A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection Sourya Dipta Das , Ayan Basak , and Saikat Dutta (https://arxiv.org/pdf/2101.03545v1.pdf arXiv:2101.03545v1 [cs.CL] 10 Jan 2021)

[17] Fake News Detection System using XLNet model with Topic Distributions: CONSTRAINT@AAAI2021 Shared Task, arXiv:2101.11425v1 [cs.CL] 12 Jan 2021 (https://arxiv.org/pdf/2101.11425v1.pdf)

[18] J. C. Reis, A. Correia, F. Murai, A. Veloso, F. Benevenuto, Supervised learning for fake news detection, IEEE Intelligent Systems 34 (2) (2019) 76–81.

[19] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, Y. Choi, Defending against neural fake news, arXiv preprint arXiv:1905.12616.

[20] Y. Bang, E. Ishii, S. Cahyawijaya, Z. Ji, P. Fung, Model generalization on covid19 fake news detection, arXiv preprint arXiv:2101.03841.

[21] K. Shu, L. Cui, S. Wang, D. Lee, H. Liu, defend: Explainable fake news detection, in: Proceedings of the 25th ACM SIGKDD International

Conference on Knowledge Discovery & Data Mining, 2019, pp. 395–405.

[22] T. Felber, Constraint 2021: Machine learning models for covid-19 fake news detection shared task, arXiv preprint arXiv:2101.03717. 32

[23] E. Shushkevich, J. Cardiff, Tudublin team at constraint@ aaai2021–covid19 fake news detection, arXiv preprint arXiv:2101.05701.

[24] O. Sharif, E. Hossain, M. M. Hoque, Combating hostility: Covid-19 fake news and hostile post detection in social media, arXiv preprint arXiv:2101.03291.

[25] A. Gautam, S. Masud, et al., Fake news detection system using xlnet model with topic distributions: Constraint@ aaai2021 shared task, arXiv preprint arXiv:2101.11425.

[26] X. Li, Y. Xia, X. Long, Z. Li, S. Li, Exploring text-transformers in aaai 2021 shared task: Covid-19 fake news detection in english, arXiv preprint arXiv:2101.02359.

[27] B. Ghanem, S. P. Ponzetto, P. Rosso, F. Rangel, Fakeflow: Fake news detection by modeling the flow of affective information, arXiv preprint arXiv:2101.09810.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, arXiv preprint arXiv:1706.03762.

[29] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805.

[30] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692.