FAKE NEWS DETECTION USING NLP AND OPTIMISED DEEP LEARNING ALGORITHMS

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE
OF

MASTER OF TECHNOLOGY IN
COMPUTER SCIENCE & ENGINEERING

Submitted by:

MITRANSHU RAJ
2K20/CSE/12

Under the supervision of
Mr. Nipun Bansal
(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering) Bawana Road, Delhi-110042

MAY, 2022

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering) Bawana Road, Delhi - 110042

CANDIDATE'S DECLARATION

I, **Mitranshu Raj**, M. Tech (Computer Science and Engineering) student Roll No. 2K20/CSE/12, hereby declare that the project Dissertation titled "**FAKE NEWS DETECTION USING NLP AND OPTIMISED DEEP LEARNING ALGORITHMS**" which I submitted to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and has not been This work has never before been used to give a degree, diploma, associateship, fellowship, or other equivalent title or honour.

Place: Delhi

Date:

Mitranshu Raj
2K20/CSE/12

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering) Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that **Mitranshu Raj**, 2K20/CSE/12, Department of Computer Science & Engineering, Delhi Technological University, Delhi, has completed a Project Dissertation titled "**FAKE NEWS DETECTION USING NLP AND OPTIMISED DEEP LEARNING ALGORITHMS**" in partial fulfilment of the requirement for the master 's degree program of Master of Technology, is an account of the project work that is carried out by the students under my supervision. This work has not been submitted in part or in full for any degree or diploma at this University or elsewhere, to the best of my knowledge.

Place: Delhi

Date:

Mr. Nipun Bansal
Assistant Professor
Department of CSE, DTU

ACKNOWLEDGMENT

The accomplishment of this project relies on the support and contribution of a large number of people as well as the institution. I am thankful to everybody who contributed to the successful project.

I'd want to show my gratitude to **Mr. Nipun Bansal**, my project guide, for allowing me to work on this project under his supervision. His unwavering encouragement and support have taught me that the process of learning is more important than the end result. Throughout all of the feedback sessions, I am appreciative to the faculty for his assistance, ongoing monitoring, and motivation to accomplish my project. He assisted me with fresh ideas, gave crucial information, and encouraged me to finish the assignment.

I also thank all my fellow students and my family for their continued support.

Mitranshu Raj
2K20/CSE/12

ABSTRACT

In today's ever-growing internet, information and news spreads rapidly like wildfire. A fraction of this news is fake or misleading, mostly for political purposes. These fake news need to be monitored. Since the amount of information generated everyday is gigantic, it is only viable to automate the fake news detection procedure. In this paper, we have thoroughly gone through various deep learning and neural network techniques that can be applicable for fake news detection. We have compared these algorithms based on there architecture and mode of operation in terms of the problem statement.

Detecting fake news is an essential yet one of the most challenging task to be be done in Natural Language Processing and Machine Learning. The tremendous rise of social media platforms has enhanced the spread of fake news while also dramatically expanding the amount of information available. As a result, false news' effect has already expanded, sometimes spilling over into media platforms and endangering public safety. Given the large volume of Web content, automatic false news detection is indeed a practical NLP task useful with any and all online copyright holders in ways that minimize work and attention in detecting and restrict the development of misinformation. The challenges of identifying fake news, and maybe even associated tasks, are discussed in this article. In this study, we offer a methodology for classifying fake news. An ensemble model comprised of pre-trained transformer-based models was utilised.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1. NLP: Natural Language Processing
2. DL: Deep Learning
3. CNN: Convolutional Neural Network
4. MNN: Multimodal attentional Neural Network
5. BERT: Bidirectional Encoder Representations for Transformers
6. LSTM: Long Short-Term Memory Network
7. SVM: Support Vector Machine
8. GAT: Graph Attention Network
9. ASSO-OSIW: Adaptive Salp swarm optimization
10. MTMN: Multimodal Topic Memory Network
11. TF-IDF: Term Frequency-Inverse Document Frequency

CHAPTER 1: INTRODUCTION

1.1 NATURAL LANGUAGE PROCESSING

Humans can transmit information with each other because utilizing natural language. They are able to converse, transmit ideas, and comprehend one other's perspectives, while the machine fails to act. To perform, the machine need commands in a timely manner. To allow computers and machines to communicate, we need to equip students to better understand natural language. Natural language processing plays an essential role in the issuance of said procedures.

Natural Language Processing (NLP) seems to be an area of computational linguistics that makes it easier humans and machines to talk directly to each other. It generally includes a friendship respectively natural & machine technologies. It goes on to tell the computer or tablet how to interact with humans through written or spoken language. It simply allows algorithms to read, analyse, and generate meaning from particular intelligence. Chatbots, interpretation, and transmission have all been component of evolution language processing.

1.2 FAKE NEWS

The title "fake news" pertains to "news articles that seem to be consciously and indisputably erroneous" in order to placate people's perspectives of certainty, events, and pronouncements. It's about resources offered as newsworthy that the promoter knows to be misleading, based on patently unsatisfactory facts or declarations or events which somehow approx. happen.

Fake news is "manufactured material which will have the presentation of main stream media television programming but without the broadcast industry's administrative guidelines and procedures for guaranteeing accuracy and completeness and trustworthiness." It is akin to fake news stories (misrepresentation information) and propaganda (false or misleading information).

The definition may seem a bit vague, but it’s important. People have used the term “fake news” to mean different things.

1.3 FAKE NEWS AND SOCIAL MEDIA

We cannot consistently turn to social media, regardless of the fact its become this almost inexorable part of our culture. On social media channels, a large volume of disinformation is not a common occurrence. Every day, people learn a bunch of things on the internet, many of which are true, but the majority of whom are not. This inaccurate and defamatory link will take to hoaxes, which is mostly made up of given tales devoid of any real data, scientific research, or summaries. Those videos are fake to convince or fool the audience The phrase "fake news" refers to a way of thinking much about character of reporting itself. The number of fake news stories has risen steadily via online channels such as What's App, Facebook, YouTube, and others, since fraudulent journals and articles are distributed online faster than we can expect. In news reporting, the term "fake news" has been used in a number of ways. In the second and third quarters of 2016, countries depicted on a Web Analytics Chart did a lot of research on the term.

The term "fake news" refers to a mindset about reality that underpins unbiased reporting. The number of fake news items has increased in recent years via social platforms such as What's App, Facebook, YouTube, and others, since bogus news outlets are distributed online faster than we can expect. In news reporting, the term "fake news" should be used in a number of ways. In the second and third quarters of 2016, nationalities depicted on a Google Scholar Chart started doing research on the topic.

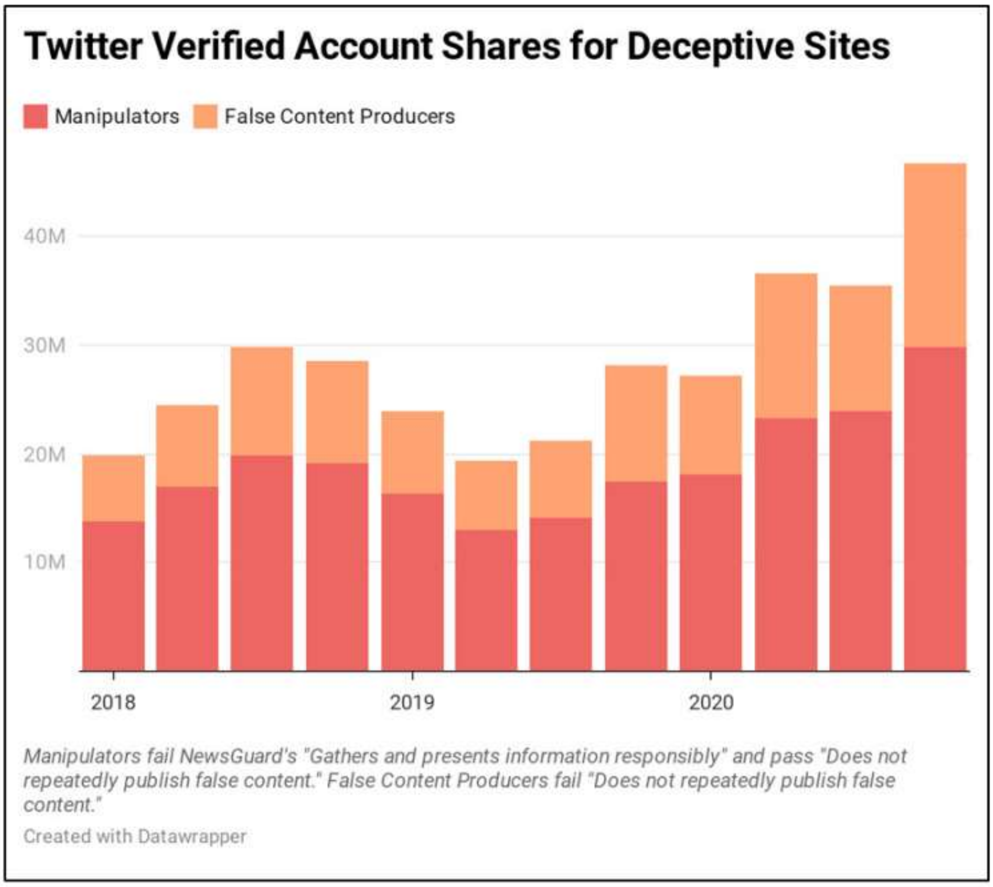


FIGURE 1 – Twitter verified account shares for Deceptive Sites

Facebook has also accomplished the commercialization of falsehoods, with manipulators boosting their engagements by 165 percent compared to the previous four years, paled in comparison to just 75 percent for where Us sites. Including all, 6.4 billion engagements were documented on deceptive websites in 2020, which is much more than two to three times number recorded four years earlier. Deceitful websites received billion interactions in the fourth quarter of 2020 alone, accounting for roughly a quarter of the entire 7 billion engagements for the us sites.

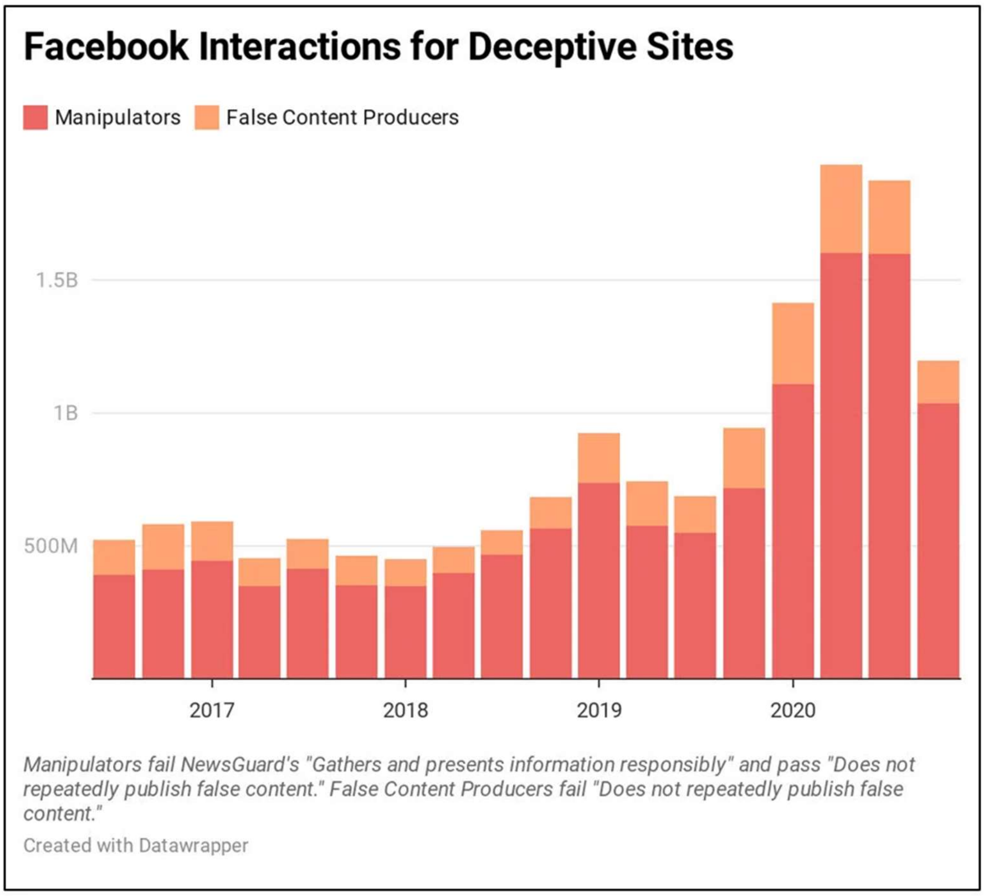


Figure 2 – Facebook Interactions for Deceptive Sites

1.4 TYPES OF FAKE NEWS

Experts in the cultural studies look at "fake news " from a range of perspectives before classifying it into distinct categories. The following sentence explains this classification:

(a) Visual-based: The material uses a graphical portrayal of picture or video edited photographs, or a combination of both, to illustrate the many sorts of false news.[1]

(b) User-based: Establishing fictional profiles that represent particular variables including such ethnicity, gender, and lifestyle might be used to attract the intended target employing this strategy. [2]

(c) Post-based: Fake news is vigorously released to the media establishments, such as Posts on facebook with movie or headshot comments, parodies, and retweets.[3]

(d) Network-based: This notion is largely utilised to organizations of connected folks on Twitter and facebook and aquaintances, where certain people in an organisation are connected to this form of bogus news. [4]

(e) Knowledge-based: To propagate fake information, these news items will be constructed utilising articles that present reasonable answers or scientific understanding about an unsolved problem. [5]

(f) Style-based: Hoaxes may be spread by all and sundry who can write in an array of designs, but this aesthetic news solely focused on how the false information being represented to legitimate consumers. [6]

Some ways for shaping whether a part of newsflash is dishonest are shown in the Figure below.

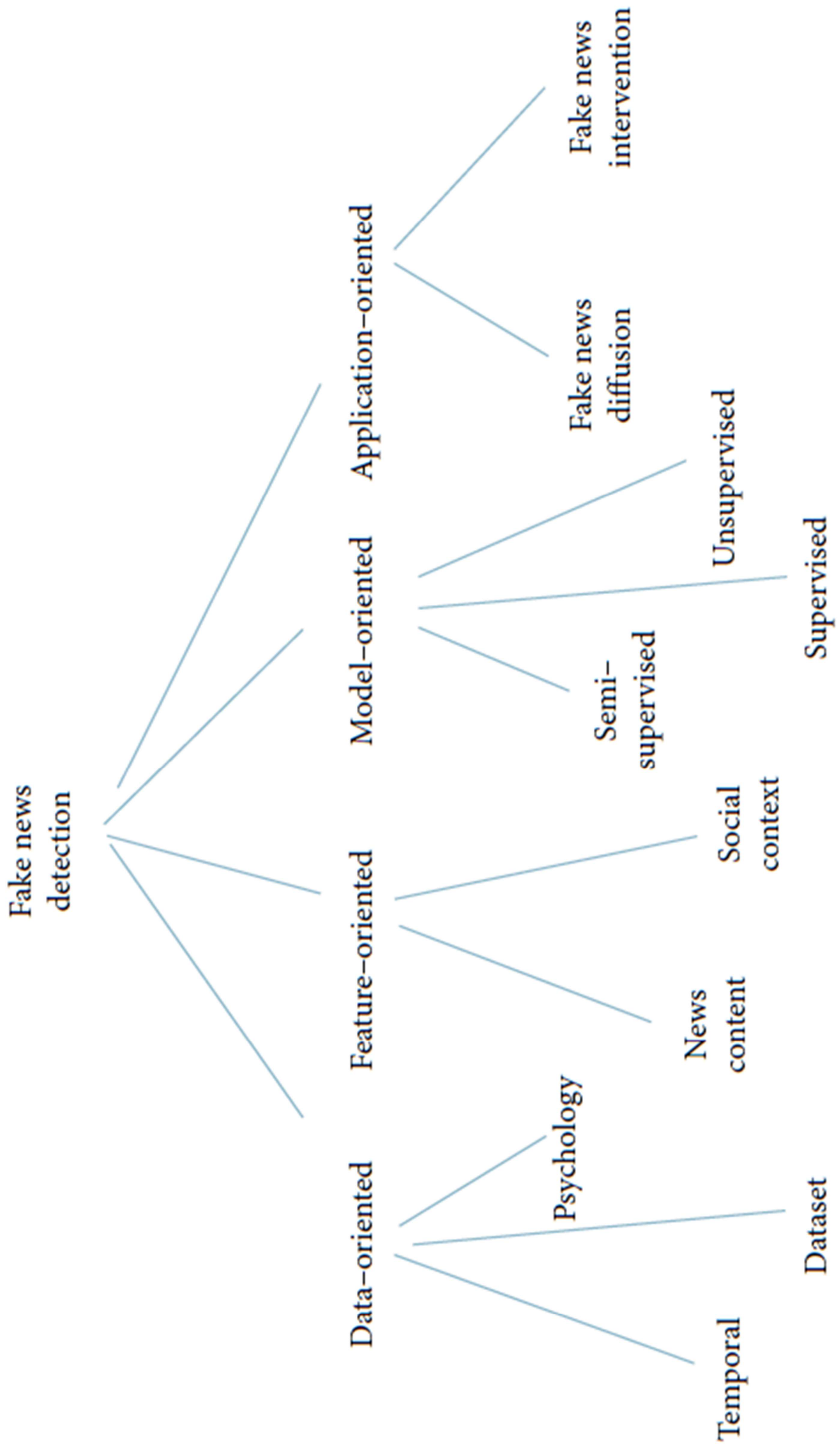


Figure 3 – Fake News Detection Techniques

1.5 Various Techniques for Identifying Fake News

In the preceding section, we explored a number of well - known methodologies for considering spoofed big stories and various types of data sets that are based on feature representation. Several sparsity techniques are described in the subsequent:

- Linguistic Feature-Based Methods. The basic language characteristics can be determined via annotation approaches. There are a plethora of other specifications, like as capitalization and extensibility, in conclusion to syntactic and quantity, with some of the most important of them were being illustrated as continues to follow:

1. n-grams: Writers accumulate words in order to identify word embedding & punctuations in a storyline. The collected attributes are saved and retrieved using TFIDF (inverse document frequency fidelity).
2. Punctuation: the FND methods use punctuation to display the dissimilarity between frank and sham texts
3. Syntax: This technique is used to extract distinctive aspects in line with both the lexico-grammatical regular expressions, which are a balance of sibling and parent vertices, as according context-free grammar (CFG). This array of attributes is then put into action.

- Deception Modeling- When it comes to telling the difference respectively bona fide and false stories, two conceptual frameworks are used: vector space modelling (VSM) and RST. RST can be used to evaluate each syntax in a hierarchical arrangement to identify rhetorical relationships. The VSM is only utilised to determine the results of the RS interactions. In contrast to similarity-based cluster analysis, the RST-VSM approach enables data edge curating given the small sample distance.

- Clustering-Based Methods. Writers have used clustering to help identify between publications that use the same clustering approach, using the Graphic Clustering Workbench clustering programme. Using clustering algorithm and also the k-nearest neighbour methodology, a large number of cliques are gauged, in addition a small amount of nodes are designed/organized, grouping equivalent newscast items based happening a processes typically of influences.

Studies have employed the idea of calculated coordinate lengths to detect if an applying concepts is false. This approach appears to be extremely effective on large datasets, according to the author's claim of a 63 percent success rate. If this approach is applied to current fake news, there is a potential that it may not produce trustworthy findings since equivalent single news sets, i.e., analogous data, might not have been sufficient.

- Non-text Cue-Based Methods: The modified version component of newsflash is used induce viewers to put implicit confidence in contaminated material, and this is the product's true motive. Numerous studies are used in this case, notably two that are categorised as:

1. Image analysis: The goal is really to alter viewers' sentiments by using a well-known way described, or smart use of visuals.
2. User behavior analysis: User behaviour analysis, a material approach, is used to assess user behaviour. The technique's main purpose is to provide a broader understanding of the social media customers' behaviour as well as the tantalizing images frequently upload.

- Content Cue-Based Methods. This approach is then developed in line with information readership' ideas and the ways wherein professionals publish news for them and for audiences. Such news pieces can be written in a variety of ways, but they mostly employ the very same core material. The following are two unique analyses presented by this technique:

1. Lexical and semantic level of analysis: readers will consider the false propaganda as true in a story as a consequence of something like the author's compelling vocabulary. To distinguish between different journalistic genres, automated algorithms can be used to retrieve the feature representation properties of text.

2. Syntactical and logical level analysis: In the conversation, the utilitarian purpose is employed to indicate the comment for subsequent portions. You may guarantee that communications are full of pointless babbling by designing enticing names.

CHAPTER 2: PRIOR WORK

Sansonetti et al. [7] conducted a comprehensive research of the features that are more prognostic for the classification of photo sharing accounts personally accountable for spreading fake news in the electronic environment, both automatically and manually. As a result, the data obtained from the watched persons includes communal and individual information, as fine as contact through content and further clienteles. An offline research employing deep learning algorithms was followed by a solid online investigation using authentic handlers to interpret reliable and untrustworthy worker accounts. The trials' discoveries, they have been scientifically confirmed, demonstration which statistics aids graphical interfaces or individuals in identifying malicious users.

Antoun et al. [8] lays out some of the most sophisticated strategies for detecting fake news in hashtags, covering algorithms for recognising bots. Unity of the recommended enlightenments just won foremost dwelling in an international competition. Models are needed to detect bogus news, and they present two here. Whenever it comes to branding, the good enough to win model takes into account differences respectively article title embeddings and the five most important matching Website search results. The new prototypical can discriminate amongst trustworthy and hoax smithereens marooned on superficial approaches to authorship, thanks to the advances in Natural Langage Understandin supervised neural examples. This new classical, developed after the tournament, significantly outperformed the prior model by a wide margin. To identify media domains, a combination strategy is used that combines named entity features with semantic embeddings produced from later part models. Time passed from when a subscription was created then the hashtag remained sent, the presence a microblog linkage, that customer's specific setting, plus related information linked with the tweet may all be second-hand to make out a Twitter bot. Some trials revealed the relative importance of various traits, and the results revealed that almost all of the recommended representations are better in rapports of efficiency.

The purpose of research remains to accurately distinguish the bulletin posture, and [9] suggestion a strategy grounded on a BERT ML model that has been pre-trained. Although some previous research categorise postures based on knowledge of only one assumption general area, this could have missed out vital info. Their approach uses multimodal conclusion to locate the optimal position because bidirectional viewpoint data can be utilised to categorise the stance quite accurately. They also categorise the stance recognition challenge as a tall structure challenge that uses subject knowledge to help classify the user's position. The outcomes of our tests show that the system seems to be more accurate in predicting a broad range of factors.

These researchers [10] devised a support vector learning system for detecting hoaxes before it reaches the general audience. They use a supervised learning differentiated instruction to deal with the huge amount of unsupervised learning on online networks. To begin, they developed the technology to capture consumer feedback, and then utilised the CredRank approach to assess the users' trustworthiness. Finally, they built a small team of connections that were aggressively disseminating a certain piece of information. SSLNews, their national press classifier, takes the outputs of those kind of three phases as inputs. SSLNews is made up of three CNN networks: one shared network, one unchecked CNN intranet, and one controlled CNN connection. They used real-world data to put Politifact and Gossipcop toward the assessment. With 25% of tagged data, the

SSLNews yielded an accuracy of 72.25 percent on Factcheck.org and 70.35 percent on Gossipcop. When using data obtained during the first 10 minutes afterwards the news broke, SSSLNews consumes a Politifact consistency of 70% in addition a Gossipcop consistency of 68% basis points.

This article [11] uses a deep learning system to detect fake news. News stories are preprocessed using a number of training prototypes before being analysed. To accomplish so, a novel transfer learning prototype has been constructed that integrates three general models, notably implementing CNN. The collaborative representation's ideal masses are derived by means of the Identity Harmony Search approach to improve fake news detection accuracy. As evidenced by the quality of the evidence we ran, the case diagram surpasses existing methodologies by 99.4 percent. Furthermore, they investigate merge incoherence and get a good accuracy of 72.3 percent.

Yanagi et al. [12] conducted a study to accurately identify wrong info premature in its dissemination, and the academics created a counterfeit information indicator that can generate imaginary community settings. The artificial context is created using fake news generators. This classifier is trained to make comments using a sample of news articles and the social milieu in which they occurred. In addition, they created a categorization model. This initiative combines news items, real-time commentary, and user-generated discussions. They tested their detector's effectiveness by analyzing the result of notes created via the identifying perfect to statements obtained authors with genuine answers. As a result, they may argue that investigating a fake remark is more effective than analysing actual comments in spotting false news. As a consequence, it appears likely our tester will be effective in detecting bogus news on popular media platforms such as Facebook and Twitter.

Paixo et al. [13] used supervised and deep learning to perform experimental investigation taking place the Fake.Br corpus, an alternative facts date in Brazilian. For the uncovering of bogus news, they provide a arrangement strategy grounded on many types of attributes in addition trained unfathomable learning models. Their top classification model achieved F1 scores of up to 95% when compared to other nondeep learning classifiers. In addition, they use uni- & bigrams to do topic modelling on a dataset to provide a complementary study.

Ren and Zhang [14] are two authors who have collaborated on a book. Hierarchical graphing attentiveness network is a reference model aimed at spam uncovering that uses a novel centralised attention approach to learn node descriptions before recognising false information. This model outperforms text-based frameworks and other network-based algorithms on two main stream media datasets. Their graph characterization knowledge or other constituent categories apps in heterogeneous networks may be improved and generalised as a consequence of testing.

This study [15] proposes a Multimodal Topic Memory Network (MTMN) that aggregates and integrates post summaries maintained across latent topics utilising global characteristics of discriminative features and represents intramodality and transformation information in a single framework.

(1) In real-world circumstances, their method incorporates a matter storage device to consciously express the final summary as a thread feature scattered between some of the persons and global properties of latent articles when freshly arriving posts contain diverse subjects scattered beyond training data. Combining two sorts of attributes yields a gathering momentum.

(2) They present a novel blended attentiveness module for simultaneous fusion that may accomplish a high degree of quality explanation while simultaneously using inter- and intramodality interconnections within every mode, as well as the

relationship between keywords and picture regions. MTMN outperforms other province algorithms in a midterm exam on two public benchmark functions.

Fake news may be recognised using both text-based approaches such as the hierarchical attention network and graphical picture characteristics like image captioning also forensic analysis, according to Meel and Vishwakarma [48]. Instead of relying just on one type of multimodal data gathering, we created captions and leads that replicated the journalism utilising . Three datasets were utilised to test each of these strategies independently before using the maximum vote ensembles process to combine them. On the hoax samples dataset, available evidence and interpretations with existing algorithms reveal that the proposed forecasting is the process the present state-of-the-art with the optimal parameters of 95.90 %.

Khan et al. [16] conducted yardstick study toward assess efficacy of several pertinent machine learning algorithms on tierce unique datasets, gathering the largest and most diversified one. They compared the performance of multiple pre - trained models linguistic concepts for identifying fake news against traditional and DL ones for its first time, and examined several aspects of the results. Pre - trained models algorithms as BERT besides perform along best occupation of recognizing dishonest news even with a small dataset. As a result, these models are an excellent alternative for languages with limited prototype material, such as testing phase. Furthermore, they conducted multiple studies based on the efficacy of the models, its subject of the article, as well as the length of the article, so they emphasised several learnings as a result of these.

Shim et al. [17] proposed a new foundation of information for sleuthing fraudulent news: the structure template of web associations usually contains news substantial. To accurately vectorize deduced composition patterns, this study proposes a new set in approach termed link2vec, which is expansion autoencoders.

Scientists tested the usefulness of their link2vec-based approach and its neutrality from vocabulary using two false news corpora in distinct languages (English and Korean). Their comparable approaches were a classic text-based sample and a cross model that combined manuscript and helps the body absorb connection information from a prior study. The link2vec-based identification prototypes were shown to be statistically significant, outperforming all other comparative concepts in mutually dialects.

Cutting-edge research by Samadi et al. [18], three classifiers with different pretrained models were recommended for embedding input news articles. Following the hidden layers, which includes new pretrained frameworks to take advantage of the deep contextualised description and recurrent convolutional categorization provided by these frameworks. Their recommended algorithms for identifying fake news beat current best practises, based on analysis from these three datasets. On LIAR, svm classifier improved by 7%, whereas on ISOT, it improved by 0.1 percent. Their team of experts produced a 1% improvement just on COVID-19 data as well.

S. Mhatre and Masurkar [19] created a strategy to detecting bogus news in this paper. The validity of two separate news reports has been determined using scraped data. Processing the information using NLP approaches including text extraction, particularly unique reduction, white-space withdrawal, and prevent removal is the first step. The next stage is encoding, which groups disagreements together grounded on their semantic affinity. Following lemmatization, a corpus is constructed using TF-IDF parallelization, which is then used to trained model. They recommend utilising a cosine matching score produced from topic modelling and the corpora to improve classification accuracy. To determine whether or not news is trustworthy, a number of classifiers are utilised, such regression models, silent encoder, and KNN Shrivastava et al. [20]. The silent predictor, it is the most often used classification algorithm in the classification of fake news, has had its training data improved. The second method is used by Zaamout et al. [21] to train a prototype model that decides whether the output is accurate or not using a supervised learning concept known as stacking and a vector comparison score. The second strategy has been found to enhance FND accuracy.

CHAPTER 3 :RESULTS AND OBSERVATIONS

Fake news detection is a problem statement with a lot of research already been conducted. With that in mind, we went through various methodologies proposed each conducted on a different dataset, with either a deep learning model implemented or an ensemble model of machine learning or deep learning model or both.

3.1 PolitiFact

PolitiFact is a dataset, that is one of the most popularly used in our problem statement universe. On PolitiFact dataset, Bi-LSTM model of ani et al [25] outperformed all the other models. Observations on PolitiFact dataset are given in Table1. Figure4 shows comparison among their accuracies.

Authors	Methodologies	Accuracy (%)
Zhou X et al [22]	Text-CNN	87.4
Shishah, W et al. [23]	BERT	84
Reddy, H. et al [24]	CNN	89
ani et al [25]	BI-LSTM	97.8
Sivasankar Elango et al. [26]	BERT	90
Shan, Guohou et al. [27]	SVM	92
Ozbay, et al [28]	ASSO-OSIW	90
Kaliyar, et al [29]	DL	90.4
Kaliyar, et al [30]	MNN	88.6
Jarrahi, et al [31]	CNN	97
Inan, E. et al [32]	GAT	91
Braşoveanu et al [33]	NLP	54
Zhou, Xinyi et al [22]	CNN	89

TABLE 1 – Comparison on PolitiFact Dataset

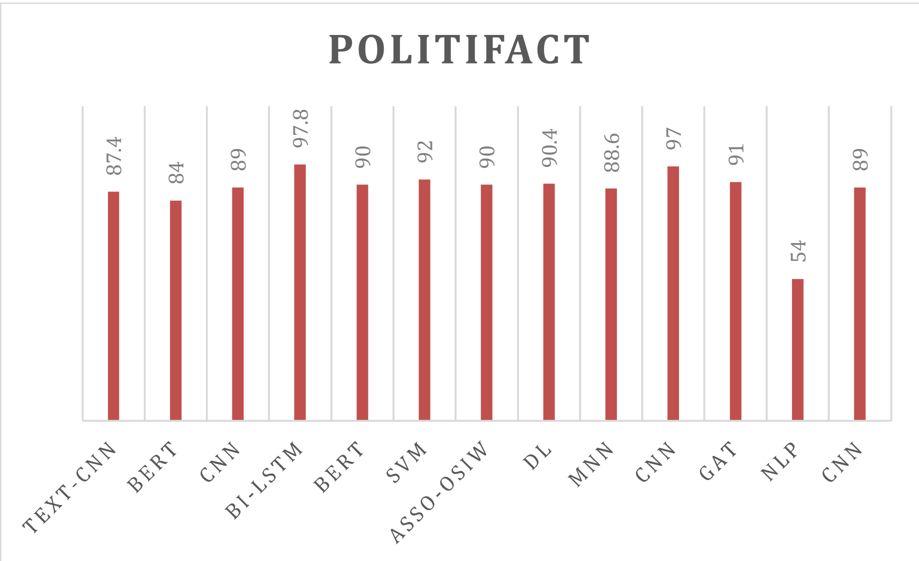


Figure 4 – Accuracy Comparison on PolitiFact Dataset

3.2 LIAR Dataset

LIAR is a dataset, that is one of the most popularly used in our problem statement universe. It has around 12k manually updated data records each with 14 attributes. On LIAR dataset, Braşoveanu, Adrian MP et al. [33] outperformed all the other models. Observations on LIAR dataset are given in Table2. Figure5 shows comparison among their accuracies.

Authors	Mehodologies	Accuracy (%)
Dong, Xishuang, et al [34]	CNN	58
Sadeghi, Fariba, et al [35]	BERT	38
Ozbay, Feyza Altunbey et al [28]	ASSO-OSIW	41
Jain, Vidit, et al [36]	CNN	46
Galli A, Masciari E,et al [37]	BERT	60
Braşoveanu, Adrian MP,et al [33]	NLP	64.3
Rajalaxmi, R. R., et al [38]	LSTM	61

TABLE 2 – Comparison on LIAR Dataset

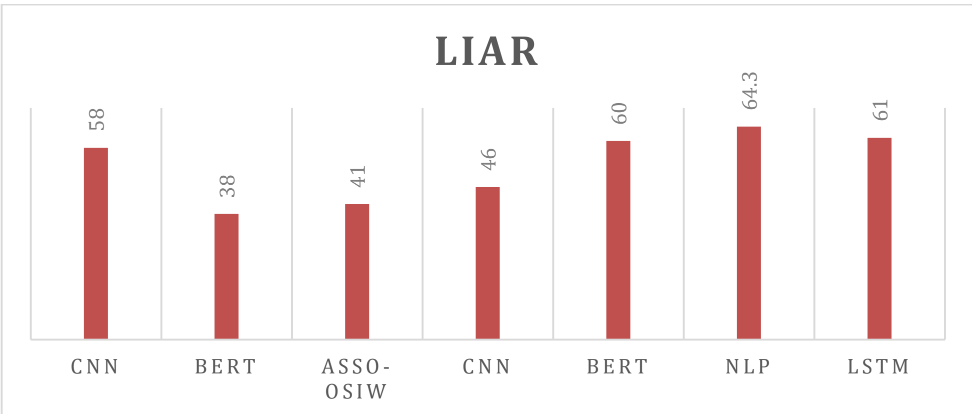


FIGURE 5 – Accuracy Comparison on LIAR Dataset

3.3 Buzzfeed

Buzzfeed is a dataset, that is one of the most popularly used in our problem statement universe. On Buzzfeed dataset, ASSO-OSIW model Ozbay, Feyza Altunbey et al. [28] outperformed all the other models. Observations on Buzzfee dataset are given in Table3. Figure6 shows comparison among their accuracies.

Authors	Methodologies	Accuracy (%)
Verma, Pawan Kumar,et al [39]	SVM	82
Ozbay, Feyza Altunbey, et al [28]	ASSO-OSIW	95
Kaliyar, Rohit Kumar, et al [29]	DL	86.36
Kaliyar, Rohit Kumar,et al [30]	MNN	86.6
Zhou, Xinyi, Atishay Jainet et al [22]	CNN	87

TABLE 3 – Comparison on Buzzfeed Dataset

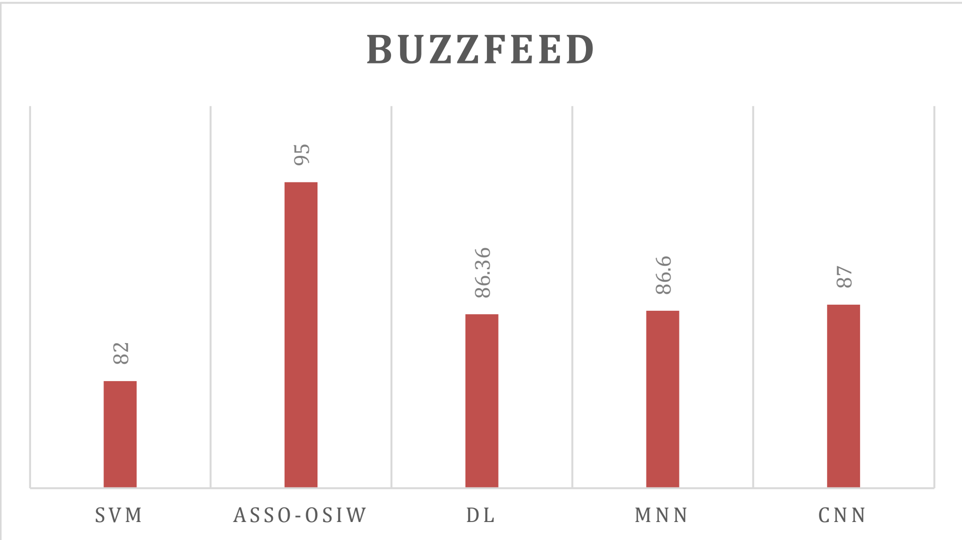


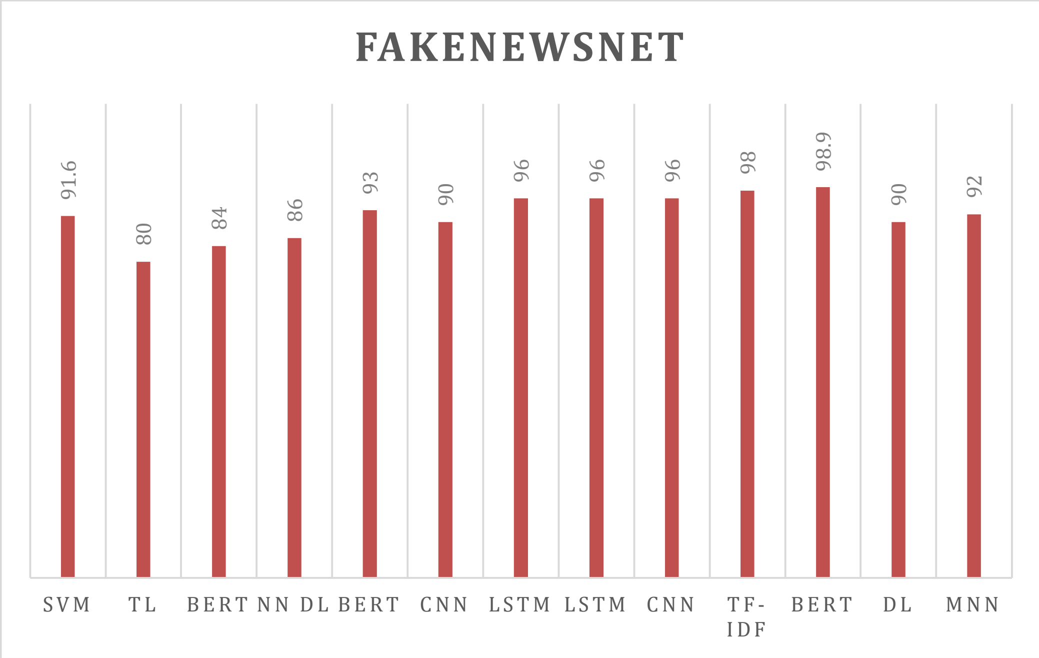
Figure 6 – Accuracy Comparison on Buzzfeed Dataset

3.4 FakeNewsNet

FakeNewsNet is a dataset, that is one of the most popularly used in our problem statement universe. On FakeNewsNet dataset, BERT model Kaliyar RK, Goswami A et al. [36] outperformed all the other models. Observations on FakeNewsNet dataset are given in Table4. Comparison among their accuracies is shown in Figure7.

Authors	Methodologies	Accuracy (%)
Verma, Pawan Kumar et al [39]	SVM	91.6
de Souza, Mariana Caravanti et al [40]	TL	80
Sadeghi, Fariba, et al [35]	BERT	84
Galende, Borja Arroyo et al [41]	NN DL	86
Nassif, Ali Bou et al [42]	BERT	93
Mughaid, Ala,et al [43]	CNN	90
Mohapatra, Asutosh et al [44]	LSTM	96
Meesad, Phayung et al [45]	LSTM	96
Malhotra, Pooja et al [46]	CNN	96
Kaur, Sawinder et al [47]	TF-IDF	98
Kaliyar RK, Goswami A et al [36]	BERT	98.9
Kaliyar, Rohit Kumar et al [30]	DL	90
Kaliyar, Rohit Kumaret al [30]	MNN	92

TABLE 4 – Comparison on FakeNewsNet Dataset



FIURE 7 – Accuracy Comparison on FakeNewsNet Dataset

3.5 PHEME

PHEME is a dataset, that is one of the most popularly used in our problem statement universe. On PHEME dataset, Ying, Long, Hui Yu et al. [48] outperformed all the other models. Observations on PHEME dataset are given in Table5. Comparison among their accuracies is shown in Figure8.

Authors	Methodologies	Accuracy (%)
Dong, Xishuang,et al [34]	CNN	59
Ying, Long, Hui Yu,et al [48]	MNN	85
Galli A, Masciari E, et al [37]	BERT	62
Ying, Long, Hui Yu, et al [48]	MN	88
Li, Ke, Bin Guo, et al [46]	GRAPHICAL MODEL	84.8
Qian, Shengsheng, et al [34]	CNN	86

TABLE 5 – Comparison on PHEME Dataset

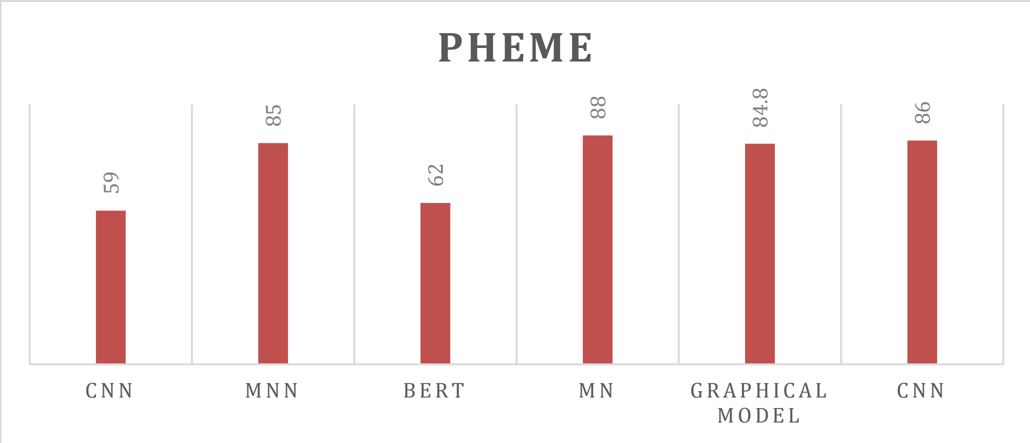


Figure 8 – Accuracy Comparison on PHEME Dataset

Chapter 4: Conclusion and Future Scope

Despite the fact that probabilistic hidden qualitative research has a high chance of success in fake news detection, rumours on social platforms is tough to pin down due to its continually changing attributes. The capacity to compute tiered features is a distinguishing feature of DL. Numerous scientific studies would therefore pertain DL methodologies like as CNNs, deep Boltzmann machines, DNN, and deep autoencoder models throughout multiple services, including audio and voice processing, NLP but rather sculpting, data information extraction, unbiased acknowledgement, and computer vision, as well as implementing DNNs, as a result of the implementation of DL practice and research in the recent times. The basic idea of hoaxes identification has indeed been detailed in full throughout this detailed evaluation, along with their categories, characteristics, and traits, as well as the taxonomies for the based on deep learning framework. Several fake sequence information have just been devised in order to identify user activity that involves making up stories or falsehoods. On several liars, false propaganda, and corpora datasets, the comparative was done for a variety of classical various computational algorithms. Deep learning approaches outscored typical ml algorithms in this examination. Bi-LSTM has the highest detection time and accuracy for bogus information throughout this assessment. This study will benefit in the introduction of unique tools or systems for rapid diagnosis of falsehoods. Additional important conclusion is that such a data may be used by the law agency's cybercrime unit, which again will aid mostly in implementation of adequate methodologies for coping with fraudulent intelligence, culminating in societal improvement. The sole drawback would be that the study is limited to corpora; but, mostly in long term, it might be extended to also include picture information in addition to textual to generate research findings over a much wider and more complex data – set.

References

[1] C. S. Atodiresei, A. Tănăselea, and A. Iftene, “Identifying fake news and fake users on Twitter,” Procedia Computer Science, vol. 126, pp. 451–461, 2018.

[2] S. Mishra, “Analysis of user sentiments using machine learning algorithms,” in 34th MP Young Scientist Congress, p. 45, MPCST, 2019.

[3] D. Bright, R. Brewer, and C. Morselli, “Using social network analysis to study crime: navigating the challenges of criminal justice records,” Social Networks, vol. 66, pp. 50–64, 2021.

[4] T. T. Aurpa, R. Sadik, and M. S. Ahmed, “Abusive Bangla comments detection on Facebook using transformer-based deep learning models,” Social Network Analysis and Mining, vol. 12, no. 1, p. 24, 2022.

[5] J. F. de Oliveira, H. T. Marques-Neto, and M. Karsai, “Measuring the effects of repeated and diversified influence mechanism for information adoption on twitter,” Social Network Analysis and Mining, vol. 12, no. 1, p. 16, 2022.

[6] H. Alzahrani, S. Acharya, P. Duverger, and N. P. Nguyen, “Contextual polarity and influence mining in online social networks,” Computational Social Networks, vol. 8, no. 1, p. 21, 2021.

[7] G. Sansonetti, F. Gasparetti, G. D’aniello, and A. Micarelli, “Unreliable users detection in social media: deep learning techniques for automatic detection,” IEEE Access, vol. 8, pp. 213154–213167, 2020.

[8] W. Antoun, F. Baly, R. Achour, A. Hussein, and H. Hajj, “State of the art models for fake news detection tasks,” in 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT), pp. 519–524, Doha, Qatar, 2020.

[9] S.-X. Lin, B.-Y. Wu, T.-H. Chou, Y.-J. Lin, and H.-Y. Kao, “Bidirectional perspective with topic information for stance detection,” in 2020 International Conference on Pervasive Artificial Intelligence (ICPAI), pp. 1–8, Taipei, Taiwan, 2020.

[10] P. M. Konkobo, R. Zhang, S. Huang, T. T. Minoungou, J. A. Ouedraogo, and L. Li, “A deep learning model for early detection of fake news on social media∗,” in 2020 7th International Conference on Behavioural and Social Computing (BESC), pp. 1–6, Bournemouth, United Kingdom, 2020.

[11] Y.-F. Huang and P.-H. Chen, “Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms,” Expert Systems with Applications, vol. 159, article 113584, 2020.

[12] Y. Yanagi, R. Orihara, Y. Sei, Y. Tahara, and A. Ohsuga, “Fake news detection with generated comments for news articles,” in 2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES), Reykjavík, Iceland, 2020.

[13] M. Paixão, R. Lima, and B. Espinasse, “Fake news classification and topic modeling in Brazilian Portuguese,” in 2020 IEEE/ WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 427–432, Melbourne, Australia, 2020.

[14] C. Song, N. Ning, Y. Zhang, and B. Wu, “A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks,” Information Processing & Management, vol. 58, no. 1, article 102437, 2021.

[15] Y. Ren and J. Zhang, “Fake news detection on newsoriented heterogeneous information networks through hierarchical graph attention,” in 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, Shenzhen, China, 2021.

[16] L. Ying, H. Yu, J. Wang, Y. Ji, and S. Qian, “Fake news detection via multi-modal topic memory network,” IEEE Access, vol. 9, pp. 132818–132829, 2021.

[17] P. Meel and D. K. Vishwakarma, “HAN, image captioning, and forensics ensemble multimodal fake news detection,” Information Sciences, vol. 567, pp. 23–41, 2021.

[18] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, “A benchmark study of machine learning models for online fake news detection,” Machine Learning with Applications, vol. 4, article 100032, 2021.

[19] J.-S. Shim, Y. Lee, and H. Ahn, “A link2vec-based fake news detection model using web search results,” Expert Systems with Applications, vol. 184, article 115491, 2021.

[20] M. Samadi, M. Mousavian, and S. Momtazi, “Deep contextualized text representation and learning for fake news detection,” Information Processing & Management, vol. 58, no. 6, article 102723, 2021.

[21] S. Mhatre and A. Masurkar, “A hybrid method for fake news detection using cosine similarity scores,” in 2021 International Conference on Communication information and Computing Technology (ICCICT), pp. 1–6, Mumbai, India, 2021.

[22] Zhou X, Wu J, Zafarani R. Safe: similarity-aware multi-modal fake news detection (2020). Preprint. arXiv. 2020;200304981.

[23] Shishah, W., 2021. Fake News Detection Using BERT Model with Joint Learning. Arabian Journal for Science and Engineering, 46(9), pp.9115-9127.

[24] Reddy, H., Raj, N., Gala, M. and Basava, A., 2020. Text-mining-based fake news detection using ensemble methods. International Journal of Automation and Computing, 17(2), pp.210-221.

[25] ani, P., Jain, V., Shokeen, J. et al. Blockchain-based rumor detection approach for COVID-19. J Ambient Intell Human Comput (2022). https://doi.org/10.1007/s12652-022-03900-2

[26] Palani, Balasubramanian, Sivasankar Elango, and Vignesh Viswanathan K. "CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT." Multimedia Tools and Applications (2021): 1-34.

[27] Shan, Guohou, Boxin Zhao, James R. Clavin, Haibin Zhang, and Sisi Duan. "Poligraph: Intrusion-tolerant and distributed fake news detection system." IEEE Transactions on Information Forensics and Security 17 (2021): 28-41.

[28] Ozbay, Feyza Altunbey, and Bilal Alatas. "Adaptive Salp swarm optimization algorithms with inertia weights for novel fake news detection model in online social media." Multimedia Tools and Applications 80, no. 26 (2021): 34333-34357.

[29] Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. "EchoFakeD: improving fake news detection in social media with an efficient deep neural network." Neural computing and applications 33, no. 14 (2021): 8597-8613.

[30] Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. "DeepFakE: improving fake news detection using tensor decomposition-based deep neural network." The Journal of Supercomputing 77, no. 2 (2021): 1015-1037.

[31] Jarrahi, A. and Safari, L., 2022. Evaluating the effectiveness of publishers' features in fake news detection on social media. Multimedia Tools and Applications, pp.1-27.

[32] Inan, E. (2022). ZoKa: a fake news detection method using edge-weighted graph attention network with transfer models. Neural Computing and Applications, 1-9.

[33] Braşoveanu, Adrian MP, and Răzvan Andonie. "Integrating machine learning techniques in semantic fake news detection." Neural Processing Letters 53, no. 5

(2021): 3055-3072.

[34] Dong, Xishuang, Uboho Victor, and Lijun Qian. "Two-path deep semisupervised learning for timely fake news detection." IEEE Transactions on Computational Social Systems 7, no. 6 (2020): 1386-1398.

[35] Sadeghi, Fariba, Amir Bidgoly, and Hossein Amirkhani. "Fake News Detection on Social Media Using A Natural Language Inference Approach." (2020).

[36] Jain, Vidit, Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, and Yashvardhan Sharma. "AENeT: an attention-enabled neural architecture for fake news detection using contextual features." Neural Computing and Applications (2021): 1-12.

[37] Galli A, Masciari E, Moscato V, Sperlí G. A comprehensive Benchmark for fake news detection. Journal of Intelligent Information Systems. 2022 Mar 21:1-25.

[38] Rajalaxmi, R. R., Narasimha Prasad, L. V., Janakiramaiah, B., Pavankumar, C. S., Neelima, N., & Sathishkumar, V. E. (2022). Optimizing Hyperparameters and Performance Analysis of LSTM Model in Detecting Fake News on Social media. Transactions on Asian and Low-Resource Language Information Processing.

[39] Verma, Pawan Kumar, Prateek Agrawal, Ivone Amorim, and Radu Prodan. "WELFake: word embedding over linguistic features for fake news detection." IEEE Transactions on Computational Social Systems 8, no. 4 (2021): 881-893.

[40] de Souza, Mariana Caravanti, Bruno Magalhães Nogueira, Rafael Geraldeli Rossi, Ricardo Marcondes Marcacini, Brucce
Neves Dos Santos, and Solange Oliveira Rezende. "A network-based positive and unlabeled learning approach for fake news detection." Machine Learning (2021): 1-35.

[41] Galende, Borja Arroyo, Gustavo Hernández-Peñaloza, Silvia Uribe, and Federico Álvarez García. "Conspiracy or Not? A Deep Learning Approach to Spot It on Twitter." IEEE Access 10 (2022): 38370-38378.

[42] Galende, Borja Arroyo, Gustavo Hernández-Peñaloza, Silvia Uribe, and Federico Álvarez García. "Conspiracy or Not? A
Deep Learning Approach to Spot It on Twitter." IEEE Access 10 (2022): 38370-38378.

[43] Galende, Borja Arroyo, Gustavo Hernández-Peñaloza, Silvia Uribe, and Federico Álvarez García. "Conspiracy or Not? A Deep Learning Approach to Spot It on Twitter." IEEE Access 10 (2022): 38370-38378.

[44] Mughaid, Ala, Shadi Al-Zu'bi, Ahmed Al Arjan, Rula Al-Amrat, Rathaa Alajmi, Raed Abu Zitar, and Laith Abualigah. "An intelligent cybersecurity system for detecting fake news in social media websites." Soft Computing (2022): 1-15.

[45] Mohapatra, Asutosh, Nithin Thota, and P. Prakasam. "Fake news detection and classification using hybrid BiLSTM and self-attention model." Multimedia Tools and Applications 81, no. 13 (2022): 18503-18519.

[46] Meesad, Phayung. "Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning." SN Computer Science 2, no. 6 (2021): 1-17.

[47] Malhotra, Pooja, and Sanjay Kumar Malik. "An Efficient Fake News Identification System Using A-SQUARE CNN Algorithm." Wireless Personal Communications (2022): 1-26.

[48] Ying, Long, Hui Yu, Jinguang Wang, Yongze Ji, and Shengsheng Qian. "Fake News Detection via Multi-Modal Topic Memory Network." IEEE Access 9 (2021): 132818-132829.

List of Publications

[1] Nipun Bansal and Mitranshu Raj, “*A Systematic Study of Fake News Detection Algorithms using Deep Learning Techniques*”, International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE), New Delhi, India, 8 May 2022.

Paper ID: IST-ICAIMLBDE-DLHI-220522-3745

INTERNATIONAL SOCIETY FOR ENGINEERING AND TECHNICAL EDUCATION

International Conference on
Artificial Intelligence, Machine Learning and Big Data Engineering

Certificate

This is to certify that Mitranshu Raj has presented a paper entitled “A Systematic Study of Fake News Detection Algorithms using Deep Learning Techniques” at the International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE) held in New Delhi, India on 08th May, 2022.

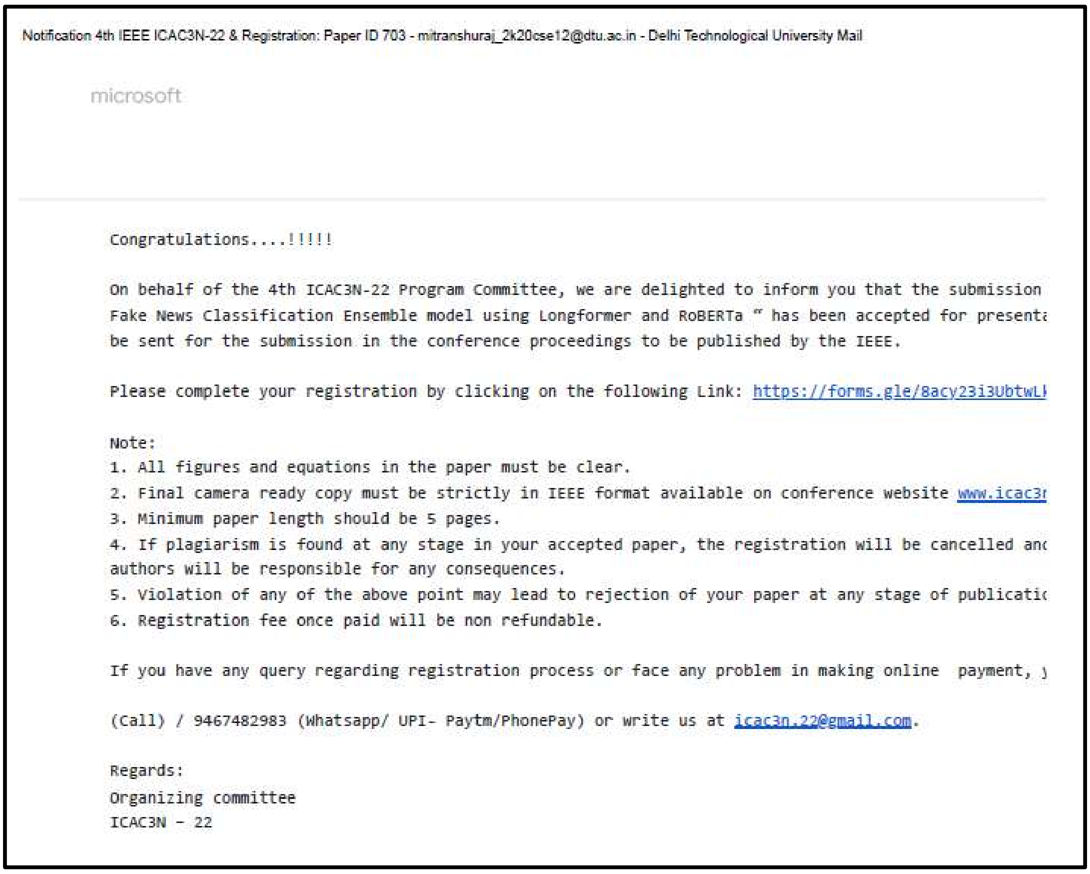
Associated with UGC Care Journals Scopus
www.isete.org | info.iseteconference@gmail.com

Conference Coordinator
International Society for Engineering and Technical Education

ISETE

Chairman
International Society for Engineering and Technical Education

[2] Nipun Bansal and Mitranshu Raj, "*A Fake News Classification Ensemble model using Longformer and RoBERTa*", 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N–22), 16th - 17th December, 2022.



Notification 4th IEEE ICAC3N-22 & Registration: Paper ID 703 - mitranshuraj_2k20cse12@dtu.ac.in - Delhi Technological University Mail

microsoft

Congratulations....!!!!!

On behalf of the 4th ICAC3N-22 Program Committee, we are delighted to inform you that the submission
Fake News Classification Ensemble model using Longformer and RoBERTa " has been accepted for presenta
be sent for the submission in the conference proceedings to be published by the IEEE.

Please complete your registration by clicking on the following Link: https://forms.gle/8acy23i3UbtwLk

Note:
1. All figures and equations in the paper must be clear.
2. Final camera ready copy must be strictly in IEEE format available on conference website www.icac3r
3. Minimum paper length should be 5 pages.
4. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled anc
authors will be responsible for any consequences.
5. Violation of any of the above point may lead to rejection of your paper at any stage of publicatic
6. Registration fee once paid will be non refundable.

If you have any query regarding registration process or face any problem in making online payment, y

(Call) / 9467482983 (Whatsapp/ UPI- Paytm/PhonePay) or write us at icac3n.22@gmail.com.

Regards:
Organizing committee
ICAC3N - 22