*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

# Handling Imbalanced Classes for Model Training in Fake News Detection

**Ashwini Deshmukh**
Department of Information Technology Department
[1]Assistant professor in Shah and anchor kutchhi engineering college ,Chembur,India
Ashwini6587@gmail.com, Orchid id:0000-0002-1441-180X

**Dr.Sharvari Govilkar**
[2]Professor and head of computer engineering Department ,Pillai's college of engineering,New Panvel ,India.
sgovilkar@mes.ac.in

Abstract—With the widespread dissemination of news on social media platforms, the propagation of fake news has become a pressing concern. Detecting fake news is crucial to maintaining the integrity of information shared across social networks. This paper presents a comprehensive investigation into the detection of fake news on social media, focusing on the collection of data from both reliable and unreliable sources To build an effective fake news detection system, a diverse dataset encompassing both reliable and unreliable sources is collected. This data collection strategy ensures a comprehensive representation of the information landscape present on social media platforms.The implementation of the bidirectional LSTM with an attention layer is a powerful approach that has shown promising results in various natural language processing tasks, including text classification and sentiment analysis. Its effectiveness lies in its ability to leverage both directional information and attention-driven focus, allowing the model to better understand and interpret the nuances of the input sequence. Calculate the class weights using the inverse of the importance factors to give proper weight to each label. Balancing has been tried with the class weight method the system has given almost 60% accuracy.

**Keywords**- *imbalanced data, class balancing, fake news, social media,Covid-19 news.*

## I. INTRODUCTION

Social media has become a ubiquitous platform that connects people from all walks of life, regardless of age or gender. It provides a space for individuals to share their thoughts, interests, and opinions with a global audience. Here are a few key points related to social media's impact on data collection, news dissemination, and the expression of diverse viewpoints:

Vast Data Collection: Social media platforms serve as a treasure trove of user-generated content, representing a wide range of interests, preferences, and behaviors. With billions of active users worldwide, social media offers an immense opportunity for data collection and analysis, providing valuable insights into user behavior, trends, and sentiments.

News Spread and Participation: Social media platforms have transformed the way news is disseminated and consumed. They enable news to spread rapidly across vast networks, reaching a large audience in a matter of minutes. Users actively engage with news content through comments, shares, likes, and retweets, fostering discussions and debates on various topics.

Expression of Opinions: Social media platforms empower individuals to freely express their opinions on a wide range of subjects. Users can engage in conversations, share their perspectives, and participate in online discussions with others who may share similar or differing viewpoints. This promotes dialogue and allows for the exchange of ideas on social, political, cultural, and other relevant issues.

Agreement and Disagreement: Social media platforms bring together diverse individuals with varying opinions and beliefs. This can lead to both agreements and disagreements on specific topics. Users have the freedom to express their agreement or disagreement through comments, tweets, and other interactive features. This creates an environment where discussions can flourish, fostering understanding and sometimes challenging prevailing narratives.

Despite the immense potential of online social media, this technology is misused to execute several undesirable acts, such as generating spam, rumors, fake messages, and fake accounts, gaining more substantial influence, creating chaos, or destabilizing homeland security. The impact of misinformation became evident during the 2016 US election, where false information was intentionally created and propagated as factual news. Social media platforms, including Twitter and Facebook, played a significant role in the spread of this propaganda. Fake accounts were created to amplify the reach of such news, exacerbating the problem. This scenario serves as a crucial case study for researchers examining the influence of social media on daily life. It highlights the importance of understanding and addressing the spread of misinformation, particularly in the context of political manipulation and propaganda [9].

Research conducted from a journalist's perspective, as outlined in reference [17], emphasizes the importance of thorough analysis and verification of news content. Journalists are trained to ask the six basic questions, known as the 5W+H (what, when, where, who, why, and how), to ensure they cover all essential aspects of a story. These questions serve as a framework to prioritize information in descending order of importance, ensuring accuracy and comprehensive reporting. detailed literature survey made helped to properly direction for our research.

## II. LITERATURE SURVEY

Early this year [3] a paper was published talking about theories and models of fake news detection. It gave a comparative analysis of the dataset used for fake news detection, the model, and the framework used. Automatic detection of fake news is not an easy task, and it has many parameters like deliberately forwarding fake news, giving biased opinions, tiring to change the thinking of a crowd through mob questioning, and so on. A comparative study of many deep learning models with its limitation. challenges and future scope were discussed. Detection of fake news continued in [4] where an exhaustive review of machine learning algorithms was done.

## III. DATASET CREATION

Collecting data on social media related to keywords like "COVID-19" and "coronavirus" can provide valuable insights into public sentiment, information dissemination, and the impact of the pandemic. Here are some approaches for data collection on social media: social media related to keywords like "COVID-19" and "coronavirus" can provide valuable insights into public sentiment, information dissemination, and the impact of the pandemic. Here are some approaches for data collection on social media:

*1. Use of API:*

Social media platforms like Twitter and Facebook provide APIs (Application Programming Interfaces) that allow developers to access public posts and retrieve data. You can utilize these APIs to gather posts, comments, and user profiles containing the desired keywords. Twitter's API, for example, provides access to real-time tweets using specific search queries.

*2. Social Media Crawlers:*

Develop or utilize existing web crawlers or scrapers specifically designed for social media platforms. These tools can navigate through public profiles, hashtags, and posts, collecting relevant data based on specified search terms.

*3. Data Mining Tools:*

utilize data mining tools or platforms that specialize in social media analytics. These platforms often provide comprehensive search and data retrieval functionalities, allowing you to collect and analyze posts, engagements, and trends related to COVID-19 and coronavirus keywords.

*4. Keyword Tracking:*

Set up keyword tracking tools to monitor social media platforms for specific keywords. These tools can track mentions, hashtags, and related discussions in real time, providing a continuous stream of data for analysis.

*5. Publicly Available Datasets:*

Explore publicly available datasets related to COVID-19 and coronavirus on platforms like Kaggle or data repositories maintained by research organizations. These datasets may contain social media data that has already been collected and processed, saving you time and effort.

*6. Surveying Social Media Users:*

Conduct surveys or questionnaires targeting social media users to gather insights, opinions, and experiences related to COVID-19. Platforms like Facebook or Reddit often have dedicated groups or forums where users actively engage in discussions about the pandemic.

When collecting data, it's important to consider the ethical aspects and ensure compliance with the terms of service and privacy regulations of social media platforms. Additionally, be mindful of the limitations and biases inherent in social media data, such as the possibility of fake accounts, echo chambers, or incomplete representations of the population. Careful data preprocessing and analysis are essential to draw accurate and meaningful conclusions from the collected social media data.

Collecting data from authentic sites like the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) can provide reliable and authoritative information about the pandemic. Here are some steps for collecting data from these sources. Collecting data from fact-checking websites like PolitiFact is a valuable approach to gathering information regarding the veracity of news and claims. These websites specialize in assessing the accuracy of statements made by public figures, politicians, and news sources. By leveraging the data from such sites, researchers can gain insights into the prevalence of misinformation and analyze trends in false or misleading information. Data collection from fact-checking websites typically involves scraping the relevant information from their articles or databases. This can include details such as the statement being fact-checked, the source of the statement, the rating given to the statement (e.g., true, false, partially true), and the supporting evidence, Researchers can utilize this data to analyze the types of claims that are commonly fact-checked, identify recurring sources of misinformation, and assess the overall accuracy of certain individuals, organizations, or news outlets. By examining the data collected from fact-checking websites like PolitiFact, researchers can contribute to a better understanding of the spread and impact of false information in public discourse.

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 9*
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

However, it is important to acknowledge that data collected from fact-checking websites may have certain limitations. The data may be biased toward the topics or claims that are fact-checked by these specific platforms, and the availability of data may vary across different fact-checking websites. Additionally, researchers should be mindful of any usage restrictions or terms of service when collecting data from these sources. or reasoning behind the rating. The volume of comments on news articles can often be substantial compared to the number of actual news articles. Comments provide an avenue for individuals to express their opinions, engage in discussions, and provide additional perspectives on the news content. While news articles themselves may be limited in quantity, analyzing the comments can offer valuable insights into public sentiment, reactions, and discussions surrounding the news.

We had a very small data collected as it was keyword specific and we wanted to collect data from reliable sites WHO and UNICEF, which don't give a huge volume of data but proper authentic data to have a good mix of unreliable news we collected verified data from PolitiFact and Snopes including attributes such as news content, data link, date of publishing articles, source, and author of the news is a comprehensive approach to data collection. These attributes provide crucial information that can contribute to the analysis and evaluation of the collected data. Let's explore the significance of each attribute:

1. *News Content:* The actual content of the news articles is a primary attribute to collect. It allows researchers to analyze the information presented, examine the language used, and assess the overall accuracy and quality of the news. The content forms the basis for further analysis, including fact-checking, sentiment analysis, and topic modeling.

2. *Data Link:* Including the data link or URL of the articles helps in source tracking and verification. It provides a reference point to access the source, which is important for transparency, fact-checking, and ensuring the credibility of the information collected.

3. *Date of Publishing Articles:* The publishing date of news articles is crucial for analyzing the temporal aspects of the data. It allows researchers to track the timeline of events, identify trends over time, and understand the context in which the news was published. It aids in establishing the chronology of news and assessing the relevance and timeliness of the information collected.

4. *Source*: Recording the source of the news articles helps researchers understand the origin and credibility of the information. It enables the analysis of different news outlets, their biases, and their reputations for accurate reporting. Considering a diverse range of sources enhances the comprehensiveness and reliability of the collected data.

5. *Author of News:* Identifying the author(s) of the news articles adds a layer of information for analysis. It allows researchers to evaluate the expertise, reputation, and potential biases of the authors. Understanding authorship provides insights into the perspectives and motivations behind news content.

By including these attributes during data collection, researchers can conduct comprehensive analyses, such as tracking the spread of information, assessing the credibility of sources, identifying patterns in news publishing, and evaluating the influence of authors on the content. These attributes form a solid foundation for further investigations into the reliability, bias, and impact of the collected data.ct checking websites.

## IV. Concept of Class Balancing

Class balancing is an important concept in real-world data collection, especially when dealing with imbalanced datasets. In many real-world scenarios, the distribution of classes or labels in the data may be highly skewed, with one class being dominant while others are underrepresented. This class imbalance can pose challenges in developing effective machine learning models, as the models tend to be biased towards the majority class and struggle to accurately predict the minority class [20],[21]
To address this issue, class balancing techniques are employed during data collection to ensure a more equitable representation of all classes. Here are a few commonly used approaches.
1. Oversampling: Oversampling involves increasing the number of instances in the minority class by replicating or creating synthetic samples. This technique helps balance the class distribution by making the minority class more prominent during training. Popular oversampling techniques include random oversampling, SMOTE (Synthetic Minority Over-sampling Technique), and ADASYN (Adaptive Synthetic Sampling).

2. Under-sampling: Under-sampling aims to reduce the number of instances in the majority class to match the minority class. This technique randomly selects a subset of the majority class samples, discarding the excess data. While under-sampling can help balance the classes, it may result in the loss of potentially valuable information. Common under-sampling methods include random under-sampling and cluster-based under-sampling.

3. Hybrid Approaches: Hybrid approaches combine oversampling and under-sampling techniques to achieve a more balanced dataset. These methods typically involve oversampling the minority class and under sampling the majority class simultaneously to strike a balance between the two.

4. Class Weighting: Class weighting is an alternative technique that assigns different weights to the classes during model training. By assigning higher weights to the minority class and lower weights to the majority class, the model gives more importance to the minority class samples. This helps mitigate the bias towards the majority class and improves the model's ability to learn from the minority class instances.

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 9***
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

We have used the class weight balancing technique for our label balancing. Class weight balancing assigns different weights to each class during the model training process, providing a way to give more importance to the minority class and reduce the bias towards the majority class.

When using class weights, the model assigns higher weights to the minority class samples and lower weights to the majority class samples. This ensures that the model pays more attention to the minority class during training, allowing it to learn from these instances more effectively. By adjusting the loss function based on the class weights, the model can achieve a better balance in its predictions and improve its ability to accurately classify the minority class.

The advantage of using class weights is that it doesn't require oversampling or under-sampling techniques, which can potentially lead to data loss or increased computational complexity. Instead, it directly influences the learning process by adjusting the impact of each class during training.

By assigning higher weights to the minority class, the model is encouraged to focus more on correctly predicting instances from that class. This helps mitigate the bias towards the majority class and allows the model to learn from the underrepresented class more effectively. During the training process, the class weights are incorporated into the loss function of the model. The weights are used to scale the contribution of each sample's loss to the overall training loss. Samples from the minority class, which have higher weights, will have a larger impact on the model's parameter updates, leading to a more balanced learning process. In our dataset label Mostly True had the lowest entries and the label FALSE has the maximum so a higher weight of 2.6607 was assigned to the label mostly true. A snip of the code below shows the class weight assigned to different labels of the dataset.

```
[ ]  weights

     {0: 0.42620137299771166,
      1: 0.6872693726937269,
      2: 1.1053412462908012,
      3: 1.8532338308457712,
      4: 2.641843971631206,
      5: 2.6607142857142856}
```

Figure 1: Class weight assigned to labels.

In Fig1 0 to 5 are labels and the numbers show the class weight assigned to it. Label False has the lowest weight of 0.426.

## V. PREPROCESSING TASK

Data preprocessing is a crucial step in preparing a dataset.
for analysis or model training. It involves transforming and cleaning the raw data to make it suitable for further processing.
1. *Data Cleaning:* This task involves handling missing values, outliers, and noise in the dataset. Missing values can be imputed using techniques such as mean, median, or interpolation. Outliers can be detected and either removed or treated depending on the nature of the data. Noise can be reduced through techniques like smoothing or filtering.

2. *Data Transformation:* Data transformation involves converting the data into a suitable format for analysis. This may include scaling numerical features to a specific range (e.g., normalization or standardization), transforming variables to meet assumptions of statistical tests (e.g., logarithmic or exponential transformations), or encoding categorical variables into numerical representations (e.g., one-hot encoding or label encoding).

3. *Feature Selection/Extraction:* In some cases, the dataset may contain many features, and not all of them may be relevant for the analysis or model training. Feature selection techniques, such as filtering (based on statistical measures) or wrapper methods (using machine learning models), can be applied to select the most informative features. Feature extraction methods, such as Principal Component Analysis (PCA) or t-SNE, can be used to create new, lower-dimensional representations of the data.

4. *Text Preprocessing:* If your dataset contains text data, text preprocessing is essential. This may include removing punctuation, converting text to lowercase, removing stop words, handling special characters, stemming or lemmatization, and tokenization. Text preprocessing helps to standardize the text data and reduce noise, making it suitable for text mining or natural language processing tasks.

5. *Splitting into Training and Test Sets*: It is common practice to split the dataset into separate training and test sets. The training set is used to build the model, while the test set is used to evaluate its performance. This helps to assess how well the model generalizes to unseen data. The split can be done randomly or using specific strategies like stratified sampling to maintain class distribution proportions.

## VI. MODEL TRAINING

Machine learning can be used in the detection of fake news by leveraging various natural language processing (NLP) and text classification technique we have trained our model using machine learning models like naive Baye, SVM, and Decision tree in our previous work. we wished to extend our work with two changes. 1. collecting our dataset
2. Trying our hands in deep learning algorithm write the following thing in a more detailed form.

In case you want to know in detail please follow our research [1] output we got in the machine learning algorithm was good. Though a lot of research has been done on deep learning models. That made us try it.

After doing data collection from the sites mentioned earlier, we pre-processed our data. We named our data as own_dataset.
To start with we did a CNN model on our dataset. Being text data, it was obvious we didn't receive a good performance. As suggested by many Literatures also we should try RNN

(recurrent neural network) model. we performed a series of experiments changing variants of own_dataset to see how results change.
All below experiments were performed on a balanced own_dataset and also on an unbalanced own_dataset.

*Experiment no 1*: using LSTM with 4 class Label (unbalanced own_dataset)
Multi-class classification can present its own set of challenges compared to binary classification tasks. In multi-class classification, instances are classified into more than two classes, making the problem more complex. To start with we started with 4 class labels instead of 6 class labels (which is our real multi-class dataset). we performed LSTM on 4 class labels and got accuracy as mentioned in the table below.

*Experiment no 2:* using LSTM with 6 class Label (unbalanced own_dataset)
After getting some confidence for 4 class labels, we performed the same task on 6 class datasets. It was good to see how increasing the number of labels makes our model learn and train more. Result of which is shown in the table below.

*Experiment no 3:* using BI-LSTM with 6 class Label (unbalanced own_dataset)
A Bidirectional LSTM (Bi-LSTM) processes the input sequence in both forward and backward directions, allowing the model to capture contextual information from both past and future contexts. This can be beneficial for tasks that require a deep understanding of the sequence. We tried Bi-LSTM on 6 label dataset and got better output as shown in the table below.

*Experiment no 4:* using BI-LSTM with attention layer with 6 class Label (unbalanced own_dataset)
The attention mechanism works by calculating attention scores for each time step based on the hidden states of the Bi-LSTM. These scores are then normalized using a SoftMax function to obtain attention weights. Finally, the attention weights are applied to the Bi-LSTM outputs, resulting in a weighted representation that emphasizes the most relevant parts of the sequence. By incorporating the attention layer, the BiLSTM model can dynamically adapt its focus and allocate more attention to specific parts of the sequence that are deemed more important for the task at hand. The output we got for this experiment is shown in the table.

TABLE I. TABLE SHOWING RESULTS OF OWN_DATASETT

| Table Column Head | | | | |
|---|---|---|---|---|
| **Data set** | **Model** | **Exp no** | **Accuracy** | **Validation Accuracy** |
| **Own_dataset** | **Unbalanced data** | **1** | **22.16** | **26.24** |
| | | **2** | **16.25** | **25.24** |
| | | **3** | **23.12** | **29.25** |
| | | **4** | **40.56** | **58.23** |
| | **Balanced data** | **5** | **20.16** | **22.24** |
| | | **6** | **18.25** | **21.24** |
| | | **7** | **25.12** | **27.26** |
| | | **8** | **42.56** | **56.32** |

Similarly, experiments were performed on balanced data with the class weight method.
Experiment no 5: LSTM with 4 class Label (balanced own_dataset)
Experiment no 6: Bi-LSTM with 6 class Label (balanced own_dataset)
Experiment no 7: Bi-LSTM with 6 class Label (balanced own_dataset)
Experiment no 8: Bi-LSTM with attention layer on 6 class Label (balanced own_dataset)

## VII. PROPOSED SOLUTION

Many datasets are available but creating our dataset was a novel method. Though few papers have done it we are happy to take it one step forward with experimenting many deep learning on it. Results may not be very promising but considering the recent literature survey we have done a great job.
We don't wish to stop here. we will try with text argument which will help us to get synonyms of words used in the dataset. We wish to increase the size of your dataset with more and more news added to it. We strongly believe in also identifying the authentication of the author along with the detection of news. We believe in working more in depth for this problem statement till we reach satisfactory conclusions.

## VIII. FUTURE SCOPE

Many datasets are available but creating our dataset was a novel method. Though few papers have done it we are happy to take it one step forward with experimenting many deep learning on it. Results may not be very promising but considering the recent literature survey we have done a great job.
We don't wish to stop here. we will try with text argument which will help us to get synonyms of words used in the dataset. We wish to increase the size of your dataset with more and more news added to it. We strongly believe in also identifying the authentication of the author along with the detection of news. We believe in working more in depth for this problem statement till we reach satisfactory conclusions.

DECLARATION

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

FUNDING

No Funds were taken for this project. All expendure was done by author on their own.

AUTHOR CONTRIBUTION

Both Authors have equally contributed to writing this research paper. AD had written the Abstract and worked on the Data creation method of the research. She is also a part of the literature survey done for this research. She has performed experiments under the guidance of second author and tabulated results.
Author SG is the guide of author AD, she has helped in finding good literature papers and helped in finding insights from them.

COMPRTING INTERESTS

We declare that the authors have no competing interests. or other interests that might be perceived to influence the results and/or discussion reported in this paper.

AVALIABILITY OF DATA AND MATERIA

The data that support the findings of this study will be available with author and can be asked through proper channel if required. We are also in process of filling copyright for our dataset.

REFERENCES

[1] A. Deshmukh and S. Govilkar, "Fake News Detection on Datasets," 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2022, pp. 274-279, doi: 10.1109/ICAST55766.2022.10039650.

[2] Lu Huang,"Deep Learning for Fake News Detection: Theories and Models",2022 Association for Computing Machinery.0 https://doi.org/10.1145/3573428.3573663,October 21–23, 2022,

[3] Nicola Capuano, Giuseppe Fenza, Vincenzo Loia , Francesco David Nota " Content-Based Fake News Detection With Machine and Deep Learning: a Systematic Review ", Neurocomputing, journal homepage: www.elsevier.com/locate/neucom,10 February 2023

[4] Ciprian-Octavian Truica˘, Elena-Simona Apostol." IT'S ALL IN THE EMBEDDING! FAKE NEWS DETECTION USING",arXiv:2304.07781v1 [cs.CL] 16 Apr 2023

[5] RamiMohawesh, SumbalMaqsood, Qutaibah Althebyan, "Multilingual deep learning framework for fake news detection using capsule neural network", Journal of Intelligent Information Systems (2023) 60:655–671 https://doi.org/10.1007/s10844-023-00788-y

[6] K. PerenArin, Deni Mazrekaj, MarcelThum, "Ability to detect and willingness to share fake news", Scientific Reports | (2023) 13:7298 | https://doi.org/10.1038/s41598-023-34402-6

[1] Aishika Pal, Pranav, Moumita Pradhan,"Survey of fake news detection using machine intelligence approach", Data and knowledge engineering,https://doi.org/10.1016/j.datak.2022.102118

[2] K. PerenArin, Deni Mazrekaj & MarcelThum," Ability to detect and willingness to share fake news", nature portfolio,https://doi.org/10.1038/s41598-023-34402-6,2023

[3] Yahya T,balquis Omar,majdi m " A Deep Learning Framework for Detection of COVID-19 Fake News on Social Media Platforms"

[4] Dasa, Ayan Basaka, Saikat Dutta "A Heuristic-driven Uncertainty based Ensemble Framework for Fake News Detection in Tweets and News Articles" arXiv:2104.01791, 5 Apr 2021

[5] Ray Oshikawa, Jing Qian, William Yang Wang "A Survey on Natural Language Processing for Fake News Detection" 2020 College of Arts and Sciences, The University of Tokyo, Department of Computer Science, University of California, Santa Barbar

[6] Ashwini Deshmukh, Garima Mahto, Jyoti Sharma , Harsh Dedhia,'Multi-Label Classification of Fake News on Social-Media ',International Journal of Innovative Science and Research Technology ,Volume 7, Issue 4, April – 2022.

[7] Michał Choraś a, Konstantinos Demestichas b, Agata Giełczyk a, Álvaro Herrero c, Paweł Ksieniewicz d, Konstantina Remoundou b, Daniel Urda c, Michał Woźniak " Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study" https://doi.org /10.1016/j.asoc.20201070 50

[8] News Detection: Survey" 2017 Albany Lab for Privacy and Security, College of Engineering and Applied Sciences University at Albany, State University of New York, Albany, NY, USA

[9] Z Khanam1, BN Alwasel1, H Sirafi1, and M Rashid2 "Fake News Detection Using Machine Learning Approaches" IOP Conf. Series: Materials Science and Engineering doi:10.1088/1757-899X/1099/1/012040

[10] Syed Ashfaq Manzoor, Dr. Jimmy Singla, Nikita "Fake News Detection using Machine Learning approaches: A systematic review" 2019 Third International Conference on Trends in Electronics and Informatics

[11] XINYI ZHOU, ATISHAY JAIN, VIR V. PHOHA, and REZA ZAFARANI, Syracuse University, USA "Fake News Early Detection: A Theory-driven Model" https://doi.org/10.1145/3377478, 2020

[12] A. Deshmukh and S. Govilkar, " Survey on fake news detection tools",ICRCSIT -Recent trends in computer science and information technology,ISBN number 978-93-80831-66-4,17-18 june 2020.

[13] Kai Shu, Xinyi Zhou Shuang Wang, Reza Zafarani, and Huan Liu "The Role of User Profiles for Fake News Detection" 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

[14] Uma Sharma, Siddharth Saran, Shankar M. Patil " Fake News Detection using Machine Learning Algorithms " International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181,2021

[15] Kyeong-Hwan Kim, Chang-sung Jeong "Fake News Detection System using Article Abstraction"2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)

[16] Tanveer Khan, Antonis Michalas, Adnan AkhunzadaSOK: Fake News Outbreak 2021: Can We Stop the Viral Spread? arXiv: 2105.10671 *22 May 2021*

[17] YounessMadaniMohammedErritaliBelaidBouikhalene"Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets"https://doi.org/10.1016 /j.rinp. 2021. 104266

[18] Kuai Xu, Feng Wang, Haiyan Wang, and Bo Yang "Detecting Fake News Over Online Social Media via Domain Reputations and Content Understanding" 2019 Tsinghua Science and Technology Huxiao Liu, Lianhai Wang, Xiaohui Han Weinan Zhang, Xun He "Detecting Fake News on Social Media: A Multi-Source Scoring 5th International Conference on Cloud Computing and Big Data Analytics, IEEE 2020

[19] Shuo Yang, Kai Shu, Suhang Wang " Unsupervised Fake News Detection on Social Media: A Generative Approach" Association for the Advancement of Artificial Intelligence,2019

[20] Ricky J. Sethi "Spotting Fake News: A Social Argumentation Framework for Scrutinizing Alternative Facts" IEEE 24th International Conference on Web Services,2017

[21] Sajjad Ahmed, Knut Hinkelmann, Flavio Corradini1 " Combining Machine Learning with Knowledge Engineering to Detect Fake News in Social Networks-a survey" Spring Symposium on Combining Machine Learning with Knowledge Engineering (AAAI-MAKE 2019). Stanford University, Palo Alto, California, USA, March 25-27, 2019.

[22] Alessandro Botticelli, Francesco Marcelloni " A survey on fake news and rumor detection techniques" Elsevier journal https: //doi. org/ 10. 1016/j.ins.2019

[23] Costel Sergiu "Identifying fake news and fake users on Twitter" International Conference on Knowledge-Based and Intelligent Information and Engineering Systems Serbia Springer 3-5 Sept 2018

[24] Hoon Ko a1, Jong Youl Hong b, Sangheon Kim c, Libor Mesicek d, In Seop Na "Human-machine interaction: A case study on fake news detection using backtracking based on a cognitive

system"www.sciencedirect.comhttps://doi.org/10.1016/j.cogsys.2018.12.
.

[25] Ghinadya, Suyanto Suyanto "Synonyms-Based Augmentation to Improve Fake News Detection using Bidirectional LSTM" 8th International Conference on Information and Communication Technology.2020

[26] Aishika Pal, Pranav, Moumita Pradhan,"Survey of fake news detection using machine intelligence approach", Data and knowledge engineering,https://doi.org/10.1016/j.datak.2022.102118

[27] K. PerenArin, Deni Mazrekaj & MarcelThum," Ability to detect and willingness to share fake news", nature portfolio,https://doi.org/10.1038/s41598-023-34402-6,2023

[28] Yahya T,balquis Omar,majdi m " A Deep Learning Framework for Detection of COVID-19 Fake News on Social Media Platforms"

[29] Dasa, Ayan Basaka, Saikat Dutta "A Heuristic-driven Uncertainty based Ensemble Framework for Fake News Detection in Tweets and News Articles" arXiv:2104.01791, 5 Apr 2021

[30] Ray Oshikawa, Jing Qian, William Yang Wang "A Survey on Natural Language Processing for Fake News Detection" 2020 College of Arts and Sciences, The University of Tokyo, Department of Computer Science, University of California, Santa B

## List of Authors

Dr. Sharvari Govilkar is a Head of Computer Engineering Department and NAAC IQAC coordinator at Pillai College of Engineering in Maharashtra, India. With over 26 years of teaching experience, she has supervised the successful completion of postgraduate studies for 38 students and is currently guiding 6 Ph.D. research scholars at Pillai College of Engineering.Dr. Govilkar is an active member of the Board of Studies for Information Technology at the University of Mumbai and is deeply committed to contributing to academics at various levels of the university. She has a significant research profile, having published 89 research papers in reputed national and international conferences and journals with a Google Scholar citation count of 782. As an expert in the areas of Data Science, Natural Language Processing, and Social Media Analytics, Dr. Govilkar is frequently invited to deliver talks on these upcoming topics in computer science at various institutions in and around
Mumbai. Her recent book on "Natural Language Processing" is an important contribution to this emerging research domain in the field of computer science and engineering. Additionally, Dr. Govilkar is an active reviewer for international conference journal andytransactions in the domain of AI and NLP

Ashwini Deshmukh is Assistant Professor in Information technology Department in Shah and Anchor kutchhi engineering college,Mumbai. She has 14 years of teaching experience. Her expertise are in the field of data science.