# Cyberspace News Prediction of Text and Image with Report Generation

.PROF.V.VINAY KRISHNA

[1]Assistant Professor,Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad,munnellimaheswari@gmail.com

[2],[3],[4]Student,Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad

ruchithagoddeti1604@gmail.com[2], sanyareddy948@gmail.com[3], bavanirayilla193@gmail.com[4]

**ABSTRACT—**The cyberspace news consumption is increasing day by day all over the world. The main reason for cyber space news consumption is due to its rapid spread of information and its easy access which lead people to consume news rapidly without the knowledge of whether the news is false or true. Thus, it leads to the wide spread of false news which leads to the negative impacts on society. Therefore, false news prediction on cyberspace is attracting a tremendous attention. The issue of fake-news prediction on cyberspace is both challenging and relevant as spreading of fake news occurs in various streams like text, audio, video, images etc. This model works on processing the text and images together by providing an interactive Application Interface (API), i.e. text by applying the model Logistic regression classifier and image by applying self-consistency algorithm. The natural language tool kit (NLTK) model is used for these implementations through python. Once the news is predicted fake, a report is redirected to the authorized website (cybercrime department) to take the immediate necessary actions required to stop this news from spreading.

## INTRODUCTION

NOWADAYS, people spend a lot of time in Internet (cyberspace) and consume news. The main reason for rapid spread of news in cyberspace is due to its low cost, easy access and easy sharing facility. This made people to consume news from cyberspace rather than fetching it from television or newspaper. The widespread of fake-news will have a serious negative impact on society and individuals.

Fake-news detection on cyberspace has led to tremendous research all over the world to predict with the exact accuracy as the content of false-news is diverse in topics. People consuming news from cyberspace produce data which is diverse and difficult to predict This model is a solution to all these problems of fake news in cyberspaces that is fast growing. In particular the datasets which are trained by various machine learning techniques like data pre-processing, feature selection, self-consistency etc. and all these are implemented by natural language processing in python.

Here we detect both forms of fake news, i.e., both text and image streams. Once the prediction is false the report is generated and it is immediately redirected to the authorized page (cybercrime department) insisting the seriousness of the news for which the actions will be taken accordingly. Through this we try to bring a safe and trustable cyberspace experience to people who rely on this. They can now verify news before they are believing or forwarding them to others.

In the age of the internet, cyberspace serves as a dynamic and ever-evolving realm where information flows incessantly, shaping our understanding of the world, influencing decision-making, and often defining the boundaries of innovation and security. The digital landscape of cyberspace is characterized by its constant flux, with news and events unfolding at breakneck speed. In this context, the ability to predict, analyse, and report on emerging news and developments within cyberspace is of paramount importance.

Cyberspace news encompasses a wide array of topics, including cybersecurity threats and breaches, technological advancements, data privacy concerns, policy changes, and the evolving digital ecosystem. Keeping abreast of these developments is not only vital for organizations and policymakers but also essential for individuals seeking to navigate the complexities of the digital age.

Traditional methods of news reporting and analysis, while valuable, often fall short in capturing the breadth and depth of the cyberspace domain. Manual monitoring of news sources and sifting through vast volumes of textual and visual data can be overwhelming, time-consuming, and may lead to information overload. Moreover, the dynamic nature of cyberspace necessitates timely and proactive reporting to mitigate risks and harness opportunities effectively.

This research embarks on a mission to address this multifaceted challenge by introducing an innovative approach: "Cyberspace News Prediction of Text and Image with Report Generation." Our aim is to develop an intelligent system that leverages the power of artificial intelligence, natural language processing (NLP), computer vision, and predictive modelling to forecast cyberspace-related news, analyse textual and image-based data, and generate comprehensive reports in real-time.

The envisioned system promises to revolutionize how we perceive, analyse, and report on cyberspace news. By seamlessly integrating textual and image data from diverse sources, our research aims to offer a holistic understanding of the cyberspace landscape. Through the deployment of advanced algorithms and predictive models, we aspire to provide insights into upcoming events, trends, and threats, allowing organizations, researchers, policymakers, and individuals to make informed decisions and take proactive measures.

In the pages that follow, we delve into the intricacies of our research methodology, the technological foundations that underpin our system, and the anticipated contributions of this study. We will explore how the fusion of textual and image data, coupled with real-time monitoring and report generation, can reshape the way we engage with cyberspace news. Additionally, we will address ethical considerations surrounding data usage and privacy, ensuring that our system adheres to responsible practices.

As we embark on this journey into the depths of cyberspace news prediction and report generation, we invite the reader to join us in exploring the potential and possibilities that lie at the intersection of artificial intelligence, information technology, and digital journalism.

Now a days the cyberspace news consumption is increasing day by day all over the world. The main reason for cyber space news consumption is due to its rapid spread of information and its easy access which lead people to consume news rapidly without the knowledge of whether the news is fake or real. It leads to the wide spread of fake news which leads to the negative impacts on society. Therefore, false news prediction on cyberspace is attracting a tremendous attention.

The main purpose of this project is to classify the news as truthful or fake using various data mining techniques. This model is a solution to all the subproblem of fake news in cyberspaces that is fast growing. In particular the datasets which are trained by various machine learning techniques like data pre-processing, feature selection, self-consistency etc. and all the sea implemented by natural language processing in python. Here we detect both forms of fake news, i.e., both text and image streams.

Once the news is Predicted as fake then the report is generated and it is immediately redirected to the authorized page (cybercrime department) insisting the seriousness of the news for which the actions will be taken accordingly. Through this we try to bring as a find trustable cyberspace experience to people who rely on this. They can now verify news before they are believing or forwarding them to others. This model works on processing the text and images together by providing an interactive Application Interface (API), i.e., text by applying the model Logistic regression classifier and image by applying self-consistency algorithm.

A. Logistic Regression Classifier logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring. An example of logistic regression could be applying machine learning to determine if a person is likely to be infected with COVID-19 or not.

B. Self-Consistency Algorithm The k-means algorithm and the principal curve algorithm are special cases of a self-consistency algorithm. A general self-consistency algorithm is described and results are provided describing the behaviour of the algorithm for theoretical distributions, in particular elliptical distributions. The results are used to contrast the behaviour of the algorithms when applied to a theoretical model and when applied to finite datasets from the model.

In this paper author is using various machine learning algorithms such as Naïve Bayes, Logistic Regression, SVM, SGD and Random Forest for fake news prediction on TEXT data and then applying VISUAL content analysis algorithm on images to predict it as fake or real news images. Most of the time fake news images are the duplicate copy of old original images so by analysing visual content of old and new images we can predict weather image uses in NEWS is real or fake.

To analyse text news author is using NLTK technique to remove stop words, special symbols, and applying stemming and lemmatization to clean text news and then converting clean text news into features vector by applying TF-IDF algorithm. TF-IDF will replace each word with its average frequency to build TF-IDF vector. TF-IDF extracted features will be input to all ML algorithms to train a model and in all algorithms Random Forest and Logistic Regression is giving better accuracy.

Cyberspace is a concept describing a widespread interconnected digital technology. "[The expression dates] from the first decade of the diffusion of the internet. It refers to the online world as a world 'apart,' as distinct from everyday reality. In cyberspace people can hide behind fake identities, as in the famous The New Yorker cartoon The term entered popular culture from science fiction and the arts but is now used by technology strategists, security professionals, governments, military and industry leaders and entrepreneurs to describe the domain of the global technology environment, commonly defined as standing for the global network of interdependent information technology infrastructures, telecommunicati ons networks and computer processing systems. Others consider cyberspace to be just a notional environment in which communication over computer networks occurs.

The word became popular in the 1990s when the use of the Internet, networking, and digital communication were all growing dramatically; the term cyberspace was able to represent the many new ideas and phenomena that were emerging.

As a social experience, individuals can interact, exchange ideas, share information, provide social support, conduct business, direct actions, create artistic media, play games, engage in political discussion, and so on, using this global network. They are sometimes referred to as cyberarts. The term cyberspace has become a conventional means to describe anything associated with the Internet and the diverse Internet culture.

The United States government recognizes the interconnected information technology and the interdependent network of information technology infrastructures operating across this medium as part of the

US national critical infrastructure. Amongst individuals on cyberspace, there is believed to be a code of shared rules and ethics mutually beneficial for all to follow, referred to as cybernetics. Many view the right to privacy as most important to a functional code of cybernetics.Such moral responsibilities go hand in hand when working online with global networks, specifically, when opinions are involved with online social experiences.

According to Chip Morningstar and F. Randall Farmer, cyberspace is defined more by the social interactions involved rather than its technical implementation.[9] In their view, the computational medium in cyberspace is an augmentation of the communication channel between real people; the core characteristic of cyberspace is that it offers an environment that consists of many participants with the ability to affect and influence each other. They derive this concept from the observation that people seek richness, complexity, and depth within a virtual world.

## RELATED WORK

### [1]Spammer Detection and Fake User Identification on Social Networks

Social networking sites engage millions of users around the world. The users' interactions with these social sites, such as Twitter and Facebook have a tremendous impact and occasionally undesirable repercussions for daily life. The prominent social networking sites have turned into a target platform for the spammers to disperse a huge amount of irrelevant and deleterious information. Twitter, for example, has become one of the most extravagantly used platforms of all times and therefore allows an unreasonable amount of spam.

Fake users send undesired tweets to users to promote services or websites that not only affect legitimate users but also disrupt resource consumption. Moreover, the possibility of expanding invalid information to users through fake identities has increased that results in the unrolling of harmful content. Recently, the detection of spammers and identification of fake users on Twitter has become a common area of research in contemporary online social Networks (OSNs). In this paper, we perform a review of techniques used for detecting spammers on Twitter.

Moreover, a taxonomy of the Twitter spam detection approaches is presented that classifies the techniques based on their ability to detect: (i) fake content, (ii) spam based on URL, (iii) spam in trending topics, and (iv) fake users. The presented techniques are also compared based on various features, such as user features, content features, graph features, structure features, and time features. We are hopeful that the presented study will be a useful resource for researchers to find the highlights of recent developments in Twitter spam detection on a single platform

**[2]A framework for realtime spam detection in Twitter**

With the increased popularity of online social networks, spammers find these platforms easily accessible to trap users in malicious activities by posting spam messages. In this work, we have taken Twitter platform and performed spam tweets detection. To stop spammers, Google Safe Browsing and Twitter's Bootmaker tools detect and block spam tweets. These tools can block malicious links;however, they cannot protect the user in real-time as early as possible. Thus, industries and researchers have applied different approaches to make spam free social network platform.

Some of them are only based on user-based features while others are based on tweet-based features only. However, there is no comprehensive solution that can consolidate tweet's text information along with the user-based features. To solve this issue, we propose a framework which takes the user and tweet-based features along with the tweet text feature to classify the tweets. The benefit of using tweet text feature is that we can identify the spam tweets even if the spammer creates a new account which was not possible only with the user and tweet-based features.

We have evaluated our solution with four different machine learning algorithms namely - Support Vector Machine, Neural Network, Random Forest and Gradient Boosting. With Neural Network, we are able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately 18%. In the past few years, online social networks like Facebook and Twitter have become increasingly prevailing platforms which are integral part of people’s daily life. People spend lot of time in microblogging websites to post their messages, share their ideas and make friends around the world. Due to this growing trend, these platforms attract a large number of users as well as spammers to broadcast their messages to the world.

Twitter is rated as the most popular social network among teenagers [2]. However, exponential growth of Twitter also invites more unsolicited activities on this platform. Nowadays, 200 million users generate 400 million new tweets per day [3]. This rapid expansion of Twitter platform influences a greater number of spammers to generate spam tweets which contain malicious links that direct a user to external sites containing malware downloads, phishing, drug sales, or scams [4]. These types of attacks not only interfere with the user experience but also damage the whole internet which may also possibly cause temporary shutdown of internet services all over the world

**[3] Enhanced Edge Preserving Restoration for 3D Images Using Histogram Equalization Technique**

An accurate and efficient face recognition system is a more interesting topic in most industries and research areas. It is a type of biometric information process that is easily adaptable as compared to the tradition card recognition system. Generally, a face recognition system is preceded by a face detection technique. The face detection technique is the preliminary stage to detect a face in live images.

In this paper, some face detection techniques are discussed such as finding skin likelihood image, skin segmentation, the morphological operation for extracting boundary regions, Haarlike features, and Ada-boost algorithm. This Haar-like feature algorithm continually searches its pattern from the particular face and which has better advantages over other techniques. After the face detection technique, the face recognition technology is applied on the detected face for further identification by using some classifiers.

**[4] Prominent features of rumour propagation in online social media**

The problem of identifying rumour's is of practical importance especially in online social networks, since information can diffuse more rapidly and widely than the offline counterpart. In this paper, we identify characteristics of rumour's by examining the following three aspects of diffusion: temporal, structural, and linguistic. For the temporal characteristics, we propose a new periodic time series model that considers daily and external shock cycles, where the model demonstrates that rumour likely have fluctuations over time.

We also identify key structural and linguistic differences in the spread of rumours and non-rumours. Our selected features classify rumours with high precision and recall in the range of 87% to 92%, that is higher than other states of the arts on rumour classification. Social psychology literature defines a rumour as a story or a statement in general circulation without confirmation or certainty to facts [1]. Rumours are known to arise in the context of ambiguity, when the meaning of a situation is not readily apparent or when people feel an acute need for security [6]. Rumours hence are a powerful, pervasive, and persistent force affecting people and groups [5]. The spread of rumours and misinformation has been studied in the context of quantifying the credibility of a given piece of information [3] and in detecting an outbreak of misinformation [10]. With the growing popularity of online social networks and their information propagation potentials, the ability to control the type of information that propagates in the network has become ever more important.

### [5] Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques

Fake news is a phenomenon which is having a significant impact on our social life, in particular in the political world. Fake news detection is an emerging research area which is gaining interest but involved some challenges due to the limited amount of resources (i.e., datasets, published literature) available. We propose in this paper, a fake news detection model that use n-gram analysis and machine learning techniques.

We investigate and compare two different features extraction techniques and six different machine classification techniques. Experimental evaluation yields the best performance using Term Frequency-Inverted Document Frequency (TF-IDF) as feature extraction technique, and Linear Support Vector Machine (LSVM) as a classifier, with an accuracy of 92%.

In the recent years, online content has been playing a significant role in swaying users' decisions and opinions. Opinions such as online reviews are the main source of information for e-commerce customers to help with gaining insight into the products they are planning to buy. Recently it has become apparent that opinion spam does not only exist in product reviews and customers' feedback. In fact, fake news and misleading articles is another form of opinion spam, which has gained traction. Some of the biggest sources of spreading fake news or rumours are social media websites such as Google Plus, Facebook, Twitters, and another social media outlet.

Even though the problem of fake news is not a new issue, detecting fake news is believed to be a complex task given that humans tend to believe misleading information and the lack of control of the spread of fake content. Fake news has been getting more attention in the last couple of years, especially since the US election in 2016. It is tough for humans to detect fake news. It can be argued that the only way for a person to manually identify fake news is to have a vast knowledge of the covered topic. Even with the knowledge, it is considerably hard to successfully identify if the information in the article is real or fake

**[6] False rumours detection on sina Wiebe by propagation structures**

This paper studies the problem of automatic detection of false rumour's on Sina Weibo, the popular Chinese microblogging social network. Traditional feature-based approaches extract features from the false rumour message, its author, as well as the statistics of its responses to form a flat feature vector. This ignores the propagation structure of the messages and has not achieved very good results.

We propose a graph-kernel based hybrid SVM classifier which captures the high-order propagation patterns in addition to semantic features such as topics and sentiments. The new model achieves a classification accuracy of 91.3% on randomly selected Weibo dataset, significantly higher than state-of-the-art approaches. Moreover, our approach can be applied at the early stage of rumour propagation and is 88% confident in detecting an average false rumour just 24 hours after the initial broadcast.

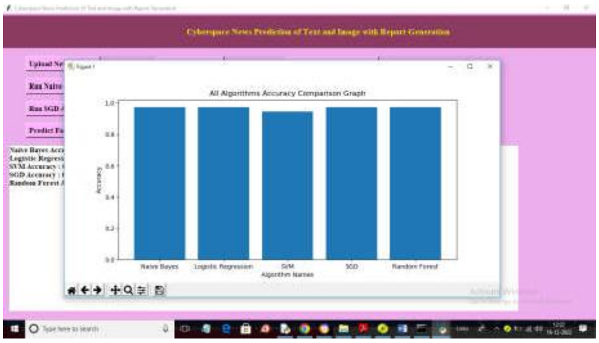
## METHODOLOGY

1) **Upload News Dataset:** using this module we will upload dataset directory with REAL and FAKE news files to application
2) **Pre-process Dataset:** using this module we will read all text news and then apply NLTK technique to clean text news data
3) **TF-IDF Vector Generation:** cleaned text news will be input to TF-IDF algorithm to convert text data into numeric vector and then convert vector into train and test

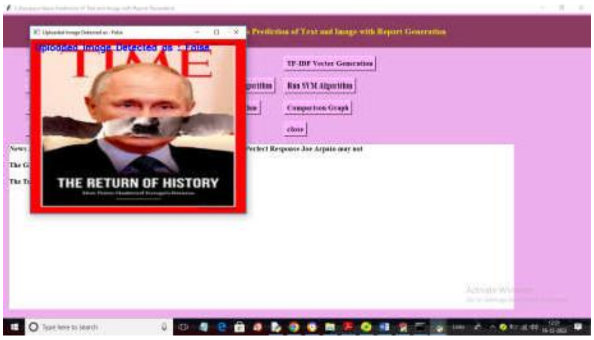
where application used 80% dataset for training and 20% for testing

4) **Run Naive Bayes Algorithm:** 80% train data will be input to Naïve Bayes algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy. The higher the accuracy the better is the algorithm
5) **Run Logistic Regression Algorithm:** 80% train data will be input to Logistic Regression algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
6) **Run SVM Algorithm:** 80% train data will be input to SVM algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
7) **Run SGD Algorithm:** 80% train data will be input to SGD algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
8) **Run Random Forest Algorithm:** 80% train data will be input to Random Forest algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
9) **Comparison Graph:** using this module we will plot accuracy graph of all algorithms
10) **Predict Fake Text News:** using this module user can enter his text and then ML algorithm will predict weather given news in TRUE or FALSE
11) **Predict Fake Image News:** using this module we will upload IMAGES and then visual analysis algorithm will predict weather image is TRUE or FALSE

## RESULT AND DISCUSSION



In above graph x-axis represents algorithm names and y-axis represents accuracy of the algorithms. Now close above graph and then click on 'Predict Fake Text News' button to enter text news and get prediction output



Uploaded Image Detected as : False

Perfect Response Joe Arpaio may not

In above screen image predicted as False News and similarly you can upload and test other images

## CONCLUSION

The consumption of news is increasing day by day in cyberspace than the traditional media. Due to its increasing popularity and user-friendly access it leaves a huge impact on individuals and society. Therefore, in this model we have found a way to detect such fake news in both the forms of text and image by using the Logistic regression model. By redirecting the fake news to the authorized website (cybercrime department), we hereby frame a high social impact and thus it reduces the spreading of false news distinctly. This model can be further discussed for the future improvement in fake news detection which can be in audio, video streams and commercialize the field to other applications.

## REFERENCES

[1] Faiza Masood, Ghana Ammad, Ahmad Almogren, Assad Abbas, Hasan Ali Khattak, Ikram Ud Din, Mohsen Guizani and Mansour Zuair, "Spammer Detection and Fake User Identification on Social Networks," IEEE Trans. Inf. Translations and content mining, vol. 7, pp. 2169- 3536, 2019.

[2] Himank Gupta, Mohd. Saalim Jamal, Sreekanth Madisetty and Maunendra Sankar Desarkar, "A framework for realtime spam detection in Twitter," IEEE Int. Conf. Communication Systems and networks, pp. 2155-2509, 2018.

[3] K. Sakthidasan, G. Srinath, Nagarajan (FEB 2014), "Enhanced Edge Preserving Restoration for 3D Images Using Histogram Equalization Technique", International Journal of Electronic Communications Engineering Advanced Research, Vol.2, SP-1, Feb.2014, pp. 40-44

[4] S. Kwon, M. Cha, K. Jung, W. Chen and Y. Wang, "Prominent features of rumor propagation in online social media," IEEE Int. Conf. Data Mining, pp. 1103–1108, 2013.

[5] Hadeer Ahmed, Issa Traore and Sherif Saad, "Detection of Online Fake News

Using N-Gram Analysis and Machine Learning Techniques," Springer, pp. 127–138, 2017.

[6] K. Wu, S. Yang, and K. Q, "False rumors detection on sina weibo by propagation structures," IEEE Int. Conf. Data Engineering, 2015.

[7] S. Sun, H. Liu, J. He, and X. Du, "Detecting event rumors on sina weibo automatically," Web Technologies and Applications, Springer, pp. 120– 131, 2013.

[8] ZhubeiJinn, Juan Compounding Zhang, Jianzhis Zhou, and Qi Tian Fellow, "Novel Visual and Statistical Image Features for Microblogs News Verification," IEEE Trans. Inf. Multimedia, pp. 1520-9210, 2016.

[9] Sanjay Yadav and Saniya Shukla, "Analysis of k-Fold Cross Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," IEEE Int. Conf. Advanced Computing, 2016.

[10] Yuanfang Guo, Xiao Chün Cao, Wei Zhang and Rui Wang, "Fake Colorized Image Detection," IEEE Trans. Inf. Information forensics and security, pp. 1556-6013, 2018.