Fake News Detection on Social Media: Variant Approaches And Comparison of Machine Learning Algorithms

[1]Aditri Rastogi, [2]Akash Mittal, [3]Anushka Saxena, [4]Bhawna Suman, [5]Ms. Sarika Tyagi

[1,2,3,4]Student, [5]Assistant Professor, Dept of Computer Science, Krishna Engineering College, Mohan Nagar Ghaziabad UP, India

Abstract - The problem of fake news has been around for as long as people have communicated. We too, might have been a victim of a false rumour for once or more in our lifetime. Moreover, the fight against fake news supply social networking sites and data utilization problems is inextricable. Misinformation related to the coronavirus COVID-19 pandemic is in the form of social media messages related to home remedies that have not been verified. Fake consultancies and conspiracy were very popular during the lockdown. P.M. Narendra Modi also made an appeal to all citizens not to believe any rumours related to the pandemic. Many Researchers have given and implemented several algorithms for detection of the Fake News. In this paper, We have used our research and several past published research papers to compare the performances of different machine learning and deep learning algorithms, which provides an idea of the most practical and efficient algorithms that can be used for fake news detection.

Keywords -Fake News Detection, Comparison, Support Vector Machine, Naïve Bayes classifier, Logistic Regression.

I. INTRODUCTION

The most significant threats to the concept of logical truth are the potential for dissemination, acceptance and ruination of false news. Fake news spread has become popular, and there has been a growing combined effort by the academic community to research and develop approaches capable of analyzing, detecting and intervening in the performance of misleading content. Letting the wrong impulses takeover is a matter due to which we fall for the false news. In the present scenario, when an average person spends twenty-four hours a week online, it clearly shows that we are constantly juggling inboxes, checking alerts, and reading the headlines feeling like we do not have time at hand to spend over reading long news texts. Humans are considered social animals, and our desire for likes can supplant a latent feeling that a story appears. Political judgement leads us to sluggish thinking, but there is an even more fundamental desire at play, i.e., our innate desire for an easy answer.

The issue of false information and fake news has been all over for as long as people have communicated. In today's low-cost, extensive communications capabilities have the potential to make the problem orders of magnitude more treacherous than in the past. Advancement in technologies makes it effortless for malicious groups to construct their alternate reality, and the urge to do it is probable to be alluring. While the extent for human distinction reduces up to a random probability of more or less 54% correctness, scientific evidence has already disclosed the human vulnerability in distinguishing true from false facts. Besides, the fight against fake news makes social network and data consumption problems indissoluble. The user only wastes network and processing resources and undermines the integrity of the service provided by spreading malicious content. As a user trusts the content provided by information sources, fake news encumbers the Quality of Trust (QoT) applied to news dispensation.

In the research paper, the recent phenomenon of fake news has been discussed, possible approaches to the differentiation of real and fake news have been presented. This research paper compares several Machine Learning and Deep Learning Algorithms which are capable of saving time, cognitive effort, and efficiency when identifying false news.

II. PROBLEM STATEMENT:

Detection of fake news is a terrifying challenge because of the minute difference between real and fake news. Humans are still inefficient in distinguishing between right(true) and wrong(false) facts that pose fake news as a warning to logical truth, which is affecting us by deteriorating democracy, journalism, and reliability on governmental organizations.

Nevertheless, there are abundant confirmations that show fake news is not too laborious to detect, at least in some selected dominions like Machine Learning and Deep

 DOI : 10.35291/2454-9150.2021.0285

Learning using different approaches.We have used our research and several past published research papers to compare the performances of different machine learning and deep learning algorithms, which provides an idea of the most practical and efficient algorithms used for fake news detection.

III. POSSIBLE APPROACHES

These are the several approaches that researchers generally use for fake news detection:

1. Language Approach

A language approach is an approach that aims at the use of linguistics to detect face news by any human or software. People who spread fake news can have reasonable control over the story they are presenting, but their language style exposes them. Language approach contemplates all the words present in a sentence and letters in a word, their structure and how it fits together in a paragraph and therefore the focus is over grammar and syntax. Three main methods that contribute to the language approach are:

Bag of Words (BOW): In the bag of words, each word in the paragraph is considered an independent entity and of equal importance. Representation of individual frequencies analyzed to find the misinformation is called n-grams. This helps in identifying patterns of word use, and with the help of these patterns, misleading information can be identified. This model is not as practical as when text is converted into numerical representation, the context and the position of a word is not considered.

Semantic Analysis: Various compatibility scores are used in this approach. By comparing a profile on the topic with personal experience, truthfulness can be determined. Writers that are truthful will mention similar remarks about a topic to other genuine writers.

Deep Syntax: Probability Context-Free Grammars are used to carry out the deep syntax method using parse trees that make Context-Free Grammar analysis possible. Various syntax structures are analyzed by converting sentences to a set of re-written rules. Fake news and real news can be differentiated by comparing syntax with the known pattern structures of lies.

2. Topic-Agnostic Approach

It approaches detect fake news by considering topic-agnostic features rather than considering the content of articles. Linguistic features and web markup capabilities are utilized in this approach. Topic – agnostic features can be a large number of advertisements, longer headlines with dramatic or striking phrases, inducing emotive responses by using different text patterns from the mainstream new or the presence of an author name.

3. Machine Learning Approach

For the identification of fake news, machine learning algorithms can be used. Various types of data sets are plied to train the algorithms.The Rumor Identification Framework, that is, a machine learning approach, has been developed that validates signals of dubious posts so that a person can quickly identify fake news. The framework will alert people of posts that might not be real.

4. Knowledge-Based Approach

Machine learning and knowledge engineering integration to detect fake news is argued in recent studies. The speed at which the news spreads over social media is challenging with some of these fact-checking methods. For eg., Twitter is a microblogging platform that can cause even a tiny piece of information to spread over a large audience very quickly. In knowledge-based approach, external sources are utilized to identify if the news is fake or real before the spread thereof becomes quicker. The knowledge-Based approach can be divided into three main categories:

(1) Expert Oriented Fact-Checking

Analyzing and scrutinizing data and documents becomes necessary with expert oriented fact-checking. Expert oriented fact-checking needs professionals to manually estimate the news's rightness via research and other studies on the specific claim. Fact-checking is the process of allocating conviction to a specific element by collating the rightness of the text to another which has formerly been fact-checked.

(2) Computational Oriented Fact-Checking

Administering users with an automated fact-checking process that can identify if a specific piece of news is real or fake is the purpose of computational oriented fact-checking. Knowledge graphs and open web sources based on practical referencing to distinguish between real and fake news are examples of computational-oriented fact-checking. ClaimBuster is a tool that has been recently developed, is an example of how fact-checking can spontaneous identify false information.

(3) Crowd Sourcing Oriented Fact-Checking

Crowdsourcing gives a group of people a chance to come to a mutual decision by examining the news's accuracy. The wisdom of the crowd is responsible for the correctness of the information. For example, Kiskkit is a platform that can be utilized for crowdsourcing, where the platform allows a group of people to assess a news article's segments. The crowd moves to the next segment for estimation after a segment has been assessed until the entire news article has been evaluated. The accuracy thence has been determined by the crowd's wisdom.

5. Hybrid Approach

Three agreed-upon elements of fake news articles are: (1) The text of an article, (2) Element is a response that an article receives and (3) the source used which motivates the news article. The hybrid model is proven to increase the percentage of humans only having a 4% chance of recognizing false news if humans make a guess and simply

 DOI : 10.35291/2454-9150.2021.0285

recognizing false news 54% of the time. Social media news with machine learning and a network approach are combined to make the hybrid model effective. Identifying the probability that the news could be fake is the purpose of this model. A hybrid model called CSI (capture, score, integrate) has been developed and functions on the main elements are - Capture (the process of extracting representations of articles by using a Recurrent Neutral Network (RNN)), Score (to create a score and representation vector) & Integrate (to integrate the outputs of the capture and score resulting in vector which is classification.

IV. PROPOSED WORK

The proposed work explores different machine learning as well as deep learning algorithms. We have used our research and several past published research papers to compare the performance of different Machine Learning and Deep Learning Algorithms, which provides an idea of the most useful and efficient algorithms that can be used for fake news detection. We have taken seven different machine learning and deep learning algorithms: Deep Syntax Analysis, Semantic Analysis, Support Vector Machine (SVM), Logistic Regression, Naïve Bayes Classifier, Neural Networks(RNN and CNN) and Decision Tree.

1. Deep Syntax Analysis: Analysis of word use is often not sufficient in predicting deception. Deeper language structures have been surveyed to predict instances of deception. Deep syntax analysis is performed through Probability Context-Free Grammars (PCFG). During Deep Syntax analysis, an application may pick different words and consider them as distinct entities while they are the same. Like, Sachin Tendulkar, Tendulkar, Master blaster, they all refer to the same person in an article, but it will choose them as different entities. So, Deep Syntax Analysis alone is not efficient for finding fake news.

2 Semantic Analysis: Semantic Analysis is the process of relating phrases, clauses, sentences and paragraphs. It may be possible that the origin of fake news is by an authentic source. Some Journalists that are well known and still they publish a fake statement, and at that time, the application will still identify it as real news because of the authenticity of the person who posted that news or tweet. So, This method has so far been restricted to fake news detection.

The two algorithms, i.e. Deep Syntax Analysis and Semantic Analysis, are not suitable for fake news detection.

So, Now we have compared the remaining five popular Machine Learning and Deep Learning Algorithms which are Support Vector Machine, Naïve Bayes Classifier, Logistic Regression,Decision Tree and Neural Networks. We have read and analyzed around 50 different Conference as well as Journal papers, and we found that out of 50 Research Papers, 25-30 Papers have used at least three of these five algorithms and have given efficient results. We have compared the five chosen algorithms' performance from 10 published Research Papers that have used these algorithms and delivered excellent results.

These papers have compared the performance of these algorithms based on their accuracy obtained on different datasets after being trained on two different Vectorizers, i.e. TF-IDF Vectorizer and Count Vectorizer.

TF-IDF Vectorizer- It shows how frequent a term is in an entire document. To represent the presence of that term, it tries to assign a metric value .This weight is used to assess how essential a word is to a report in a corpus.

Count Vectorizer- It represents a record in the form of a matrix data set matrix notation in which a corpus document is represented by each row, each column represents a corpus term, and each cell shows the frequency output of a particular term in a specific document.

The comparison of performances of the algorithms based on Research Papers and different datasets is shown in the table below:

Ref.	Authors	Datasets used	Algorithms used	Accuracy (Count Vectorizer)	Accuracy (TF-IDF Vectorizer)	Highest Accuracy
[1]	Karishnu Poddar, Geraldine Bessie Amali D,Umadevi K S	13000 articles from 244 websites	SVM Naïve Bayes Logistic Regression Decision Tree Neural Networks	0.89 0.86 0.87 0.82 0.49	0.92 0.85 0.91 0.81 0.49	SVM
[2]	Abdullah-All-Tanvir, Ehesas Mia mahir,Mohammed Rezwanul Huq	20,360 tweets from Twitter	SVM Naïve Bayes Logistic Regression Decision Tree Neural Networks	0.89 0.84 0.62 0.74 0.73	0.89 0.89 0.69 0.76 0.73	SVM

 DOI : 10.35291/2454-9150.2021.0285

[3]	Smitha. N, Bharath .R	15000 posts from 213 websites	SVM Naïve Bayes Logistic Regression Long Short Term Memory(LSTM) Neural networks	0.91 0.93 0.82 0.88 0.91	0.94 0.90 0.93 0.82 0.93	SVM And Naïve Bayes
[4]	Anjali Jain, Avinash Shakya, Harsh Khatter , Amit Kumar Gupta	Data from Facebook and Twitter	SVM Naïve Bayes Logistic Regression Decision Tree Neural Networks	0.93 0.93 0.82 0.86 0.78	0.94 0.96 0.85 0.76 0.73	Naïve Bayes
[5]	Sahil Gaonkar, Sachin Itagi,Avinash Gaonkar	948,373 messages by Twitter API	SVM Naïve Bayes Logistic Regression Decision Tree Neural Networks	0.99 0.89 0.97 0.76 0.89	0.94 0.85 0.93 0.73 0.78	SVM
[6]	Arush Agarwal, Akhil Dixit	20,801 News Reports	SVM Naïve Bayes Logistic Regression Decision Tree Random Forest	0.73 0.91 0.97 0.87 0.76	0.75 0.92 0.84 0.73 0.84	Naïve Bayes
[7]	Jun Lin, Glenna Tremblay-Taylor, Guanyi Mou, Di You,Kyumin Lee	22,456 News Articles	SVM Naïve Bayes Logistic Regression Decision Tree Neural Networks	0.77 0.78 0.83 0.78 0.82	0.87 0.89 0.94 0.85 0.81	Logistic Regression
[8]	Irfan Kareem, Shahid Mahmood Awan	344 news Articles	SVM Naïve Bayes Logistic Regression Convolutional Neural Network(CNN) Recurrent Neural Networks(RNN)	0.85 0.75 0.88 0.65 0.62	0.86 0.71 0.92 0.77 0.74	Logistic Regression
[9]	Rahul R Mandical, Mamatha N, Shivakumar N, Monica R, Krishna A N	8000 News Articles	SVM Naïve Bayes Logistic Regression Decision Tree Neural Networks	0.98 0.87 0.78 0.77 0.67	0.89 0.92 0.74 0.72 0.76	SVM
[10]	Mohamed K.Elhadad, Kin Fun Li, Fayez Gebali	Data from Facebook and Twitter	SVM Naïve Bayes Logistic Regression Decision Tree Long Short Term Memory(LSTM)	0.78 0.87 0.76 0.65 0.67	0.74 0.92 0.75 0.64 0.63	Naïve Bayes

V. CONCLUSION

This research paper compares the performances of five well-Known machine learning algorithms for Fake News Detection, i.e., Support Vector Machine, Naïve Bayes, Logistic Regression, Neural Networks (Recurrent Neural Network and Convolutional Neural Network), and Decision Tree. These Machine Learning Algorithms are used in different Research papers with mostly these two Vectorizers, i.e., TF-IDF Vectorizer and Count Vectorizer.

VI. FUTURE SCOPE

This paper compared performances of five well-known machine learning algorithms, i.e., Support Vector Machine, Naïve Bayes, Logistic Regression, Neural Networks (Recurrent Neural Network and Convolutional Neural Network), and Decision Tree. We concluded that out of all The compared Machine Learning Algorithms targets to minimize time, maximize accuracy, and resource utilization. Our analysis mainly considers Datasets and Research from different Research Papers. After the observations made from the above table, i.e., highest accuracy with both TF-IDF Vectorizer and Count Vectorizer, we concluded that Support Vector Machine (SVM), Naïve Bayes and Logistic Regression are the three most suitable machine learning algorithms that can be used for Fake News Detection.

 DOI : 10.35291/2454-9150.2021.0285

the Algorithms, the following could be the most suitable ones for fake news detection:

Naive Bayes (to consider identical common properties between spam message and news articles which are fake)

Logistic Regression (Maximum Likelihood (ML) Estimation which says coefficients should be chosen in such a way that it maximizes the probability of Y given X (likelihood))

Support Vector Machine (SVM) (which will help classify the news and make it fall under a category for easy processing, will help analyze the news and give output as the news is fake or real in a more precise manner.)

The future scope of this paper will be the comparison between the performances of the three most suitable machine learning algorithms, i.e., Support Vector Machine, Logistic Regression, and Naïve Bayes for Fake News Detection. These three machine learning algorithms will be trained on the dataset with TF-IDF Vectorizer and Count Vectorizer. On the basis of the results, the accuracy will be compared, and the algorithm with the highest accuracy and best results will be considered as the best machine learning algorithm For Fake News Detection.

REFERENCES

[1]. Karishnu Poddar, Geraldine Bessie Amali D, Umadevi K S "Comparison of Various Machine Learning Models for Accurate Detection of Fake News" Innovations in Power and Advanced Computing Technology (i-PACT) -2019

[2]. Abdullah-All-Tanvir, Ehesas Mia mahir, Saima Akhter, Mohammed Rezwanul Huq "Detecting Fake News Using machine Learning and Deep Learning Algorithms" Institute of Electrical and Electronics Engineers [IEEE] Journal -2019

[3]. Smitha. N, Bharath .R , "Performance Comparison of Machine Learning Classifiers for Fake News Detection" International Conference on Inventive Research in Computing Applications (ICIRCA-2020) -2020

[4]. Anjali Jain, Avinash Shakya, Harsh Khatter, Amit Kumar Gupta "A Smart System For Fake News Detection Using Machine Learning" International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) -2019

[5]. Sahil Gaonkar, Sachin Itagi, Avinash Gaonkar "Detection Of Online Fake News : A Survey" International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) -2019

[6]. Arush Agarwal, Akhil Dixit "Fake News Detection: An Ensemble Learning Approach" International Conference on Intelligent Computing and Control Systems (ICICCS 2020) -2020

[7]. Jun Lin, Glenna Tremblay-Taylor, Guanyi Mou, Di You, Kyumin Lee "Detecting Fake News Articles" IEEE International Conference on Big Data (Big Data) -2019

[8]. Irfan Kareem, Shahid Mahmood Awan "Pakistani Media Fake News Classification using Machine Learning Classifiers" International Conference on Innovative Computing (ICIC) -2019

[9]. Rahul R Mandical, Mamatha N, Shivakumar N, Monica R, Krishna A N, "Identification of Fake News Using Machine Learning" Department of Computer Science and Engineering SJB Institution of Technology -2020

[10]. Mohamed K. Elhadad, Kin Fun Li, Fayez Gebali "A Comparison of Machine Learning Algorithms in Fake News Detection" Institute of Electrical and Electronics Engineers [IEEE] Journal -2019

[11]. Nicollas R. de Oliveira, Dianne S. V. Medeiros and Diogo M. F. Mattos "A Sensitive Stylistic Approach to Identify Fake News on Social Networking" Institute of Electrical and Electronics Engineers [IEEE] Journal -2020

[12]. Akshay Pratap Singh, Deepanshi Sachdeva and Dheeraj Kumar "Fake News Detection On Social Media" International Journal of Scientific Research in Engineering and Management[IJSREM] –July 2020

[13]. Kuai Xu , Feng Wang, Haiyan Wang, and Bo Yang "Detecting Fake News Over Online Social Media via Domain Reputations and Content Understanding" Institute of Electrical and Electronics Engineers [IEEE] Journal -2020

[14]. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang and Huan Liu "Fake News Detection on Social Media: A Data Mining Perspective" ACM SIGKDD Explorations Newsletter-2017.

 DOI : 10.35291/2454-9150.2021.0285