International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST229245

A Review of Machine Learning-Based Fake News Analysis

Vaishali Singh
Assistant Professor, Department of Computer Science and Engineering, Ambalika Institute of Management & Technology, Lucknow, Uttar Pradesh, India

Article Info
Volume 9, Issue 2
Page Number : 252-261

Publication Issue
March-April-2022

Article History
Accepted : 01 April 2022
Published : 09 April 2022

ABSTRACT

In this paper considers the use of NLP (Natural Language Processing) methods for identifying Fake news, that is, deceptive reports that come from untrustworthy sources. Simply building a model based on a tally vectorizer (using word counts) or a (Term Frequency Inverse Document Frequency) tfidf framework (word counts compared to how frequently they're used in different articles in your dataset) can only get you so far. However, these models do not take into account critical aspects like word requesting and setting. It is entirely possible that two articles that are similar in their promise include will be completely different in their significance. The information science community has reacted by taking action against the problem. There is a competition called the "Fake News Challenge," and Facebook is using AI to sift fake reports through client channels. Combating Fake News is an excellent book arrangement project with a simple recommendation. Is it possible for you to build a model that can distinguish between "Genuine" and "Fake" news? As a result, a proposed work on amassing a dataset of both fake and genuine news and using a Naive Bayes classifier to create a model to classify an article as fake or genuine based on its words and expressions.

Keywords : Naïve Bayes, Natural Language Processing (NLP), Real News, Fake News, Term Frequency Inverse Document Frequency (tfidf).

I. INTRODUCTION

Previously, if someone needed news, they would wait for the next day's newspaper. However, [1] with the development of online papers that update news in a split second, people have discovered a superior and quicker way to be educated about the issue of their choice [2]. Nowadays, social interaction frameworks, online news entrances, and other [3] online media have become the primary sources of information [4] through which fascinating and breaking news are shared at a rapid pace. Regardless, numerous news entryways serve a specific interest by dealing with twisted, somewhat correct, and once in a while fanciful news that is likely to catch the attention of a specific group of people. Counterfeit news [5] has become a major source of concern due to its potential for causing confusion and intentional disinformation among individuals [6].The term counterfeit news has become a popular expression nowadays. In any case, a concurred meaning of the expression "counterfeit news [7] is still to be found. It tends to be characterized as a sort of sensationalist reporting or purposeful publicity that comprises of intentional deception or fabrications spread through customary print and broadcast news media or online web-based media [8]. These are distributed for the most part with the aim to deceive to harm a local area or individual, make disarray, and gain monetarily or strategically. Since individuals are frequently unfit to invest sufficient energy to cross-check reference and make certain of the believability of information, robotized location of phony news is essential. Along these lines, it is getting incredible consideration from the examination local area. There are numerous examples where keenly planned phony news had extreme outcome by impelling strict or ethnic gatherings against guiltless casualties. On October 17, 2018, United States Congressman Matt Gaetz (R-FL) presented a video on [9] Twitter and proposed, without proof that showed a gathering of individuals being paid by tycoon George Soros to join a transient train and tempest the United States line. The video was miscaptioned and the tweet contained verifiable inaccuracies.1 On 23 June 2018, a progression of appalling pictures and recordings started to circle on Facebook. One showed a monitors skull hacked open that was seen in excess of multiple times. The Facebook clients who posted the pictures guaranteed they showed a slaughter in progress in the Gashish area of Plateau State, Nigeria by Fulani Muslims who were murdering Christians from the locales Berom ethnic minority. As an outcome, a slaughter occurred in Gashish that end of the week and somewhere close to 86 and 238 Berom individuals were executed, as per gauges made by the police and by neighborhood local area pioneers. Nonetheless, probably the most combustible pictures and recordings were absolutely immaterial to the brutality in Gashish. The video showing a man's head was cut, was not occurred in Nigeria and it was recorded in Congo, in 2012.2 The earlier chips away at counterfeit news location have applied a few conventional AI [10] techniques and neural organizations to distinguish counterfeit news. In any case, they have zeroed in on recognizing information on specific sorts, (for example, political) [11]. In like manner, they fostered their models and planned highlights for explicit datasets that match their subject of interest. All things considered, these methodologies would experience the ill effects of dataset predisposition and are probably going to perform ineffectively on information on another point. A portion of the current investigations have likewise made correlations among various strategies for counterfeit news recognition. It has assembled a benchmark dataset specifically, Liar and tested some current models on that dataset. The examination result hints us how various models can perform on an organized dataset like Liar. Be that as it may, the length of this dataset isn't adequate for neural organization investigation and a few models were found to experience the ill effects of overfitting. Gilda has investigated some conventional AI approaches [12]. Notwithstanding, many progressed AI models, e.g., neural organization based ones are not applied that have been demonstrated best in numerous content characterization issues. A significant limit of earlier relative examinations is that these are completed on a particular sort of dataset [13], it is hard to arrive at a decision about the exhibition of different models. Additionally, these works have zeroed in on a predetermined number of highlights that have brought about the deficient investigation of expected attributes of phony news. In this examination, we will probably introduce a relative presentation investigation of existing strategies by carrying out every one on two of the accessible datasets and another pre-arranged by us consolidating information on circulated subjects. We likewise fuse various highlights from existing works and explore the exhibition of some effective content order strategies that are yet to be applied for counterfeit news recognition as far as we could possibly know. There exists a huge assemblage of exploration on the subject of AI techniques for trickiness discovery, its vast majority has been zeroing in on ordering on the web audits and freely accessible online media posts. Especially since late 2016 during the American Presidential political race, the topic of deciding 'counterfeit news' has likewise been the subject of specific consideration inside the writing. Conroy, Rubin, and Chen [14] diagram a few methodologies that appear to be encouraging towards the point of impeccably group the deceptive articles. They note that basic substance related n-grams and shallow grammatical features (POS) labeling have demonstrated inadequate for the characterization task, frequently neglecting to represent significant setting data. Maybe, these techniques have been shown valuable just couple with more perplexing strategies for examination. Profound Syntax examination utilizing Probabilistic Context Free Grammars (PCFG) have been demonstrated to be especially important in blend with n-gram techniques. Feng, Banerjee, and Choi [15] can accomplish 85%-91% precision in trickery related arrangement assignments utilizing on the web audit corpora. Feng and Hirst carried out a semantic examination taking a gander at 'object: descriptor' sets for logical inconsistencies with the content on top of Feng's underlying profound sentence structure model for extra improvement. Rubin, Lukoianova and Tatiana examine logical construction utilizing a vector space model with comparable achievement. Ciampaglia et al. utilize language design similitude networks requiring a prior information base. In this review paper section I contain the introduction, section II contains the literature review details, section III contains the details about methodologies, section and section IV provide conclusion of this paper.

II. RELATED WORK

This section has taken into account an extensive literature survey related to fake news analysis using Machine Learning. Ethar Qawasmeh et. al. (2019) [14] fast advancement of figuring patterns, remote interchanges, and the keen gadgets industry has added to the inescapable of the web. Individuals can get to internet providers and applications from anyplace on the planet whenever. There is no uncertainty that these innovative advances have made our lives simpler and saved our time and endeavors. On the opposite side, we ought to concede that there is an abuse of web and its applications including on the web stages. For instance, online stages have been engaged with getting out counterfeit word everywhere on the world to fill certain needs (political, monetary, or web-based media). Identifying counterfeit news is viewed as one of the hard difficulties in term of the current substance based examination of customary techniques. As of late, the exhibition of neural organization models has beated conventional AI techniques because of the remarkable capacity of highlight extraction. All things considered, there is an absence of exploration work on distinguishing counterfeit news in news and time basic occasions. Along these lines, in this paper, we have examined the programmed recognizable proof of phony news over online correspondence stages. Besides, we propose a programmed ID of phony news utilizing current AI procedures. The proposed model is a bidirectional LSTM connected model that is applied on the FNC-1 dataset with 85.3% precision execution.

William Yang Wang (2018) [15] automatic phony news identification is a difficult issue in misdirection discovery, and it has huge true political and social effects. Be that as it may, measurable ways to deal with battling counterfeit news has been drastically restricted by the absence of marked benchmark datasets. In this paper, we present LIAR: another, freely accessible dataset for counterfeit news recognition. We gathered a long term, 12.8K physically marked short explanations in different settings from POLITIFACT.COM, which gives nitty gritty examination report and connections to source records for each case. This dataset can be utilized for certainty checking research too. Prominently, this new dataset is a significant degree bigger than already biggest public phony news datasets of comparable sort. Observationally, we examine programmed counterfeit news recognition [16] dependent on surface-level etymological examples. We have planned a novel, half breed convolutional neural [17] organization to incorporate metadata with text. We show that this crossover approach can improve a book just profound learning model. Z Khanam, et. al., (2021) [18] fake news via online media and different other media is wide spreading and involves genuine worry because of its capacity to cause a ton of social and public harm with ruinous effects. A great deal of exploration is as of now centered around distinguishing it. This paper makes an investigation of the examination identified with fake news discovery and investigates the conventional AI models to pick the best, to make a model of an item with directed AI [19] calculation, that can group counterfeit news as evident or bogus, by utilizing apparatuses like python scikit-learn, NLP for text based examination. This cycle will bring about highlight extraction and vectorization; we propose utilizing Python scikit-learn library to perform tokenization and highlight extraction of text information, since this library contains helpful apparatuses like Count Vectorizer and Tiff Vectorizer. Then, at that point, we will perform include determination techniques, to try and pick the best fit highlights to acquire the most elevated exactness, as indicated by disarray framework results.

Costin BUSIOC et. al., (2020) [20] fighting phony news is a troublesome and testing task. With an expanding sway on the social and world of politics, counterfeit news apply an unprecedently sensational effect on individuals' lives. Because of this marvel, drives tending to computerized counterfeit news discovery have acquired prominence, producing inescapable examination interest. Notwithstanding, most methodologies focusing on English and low-asset dialects experience issues when conceiving such arrangements. This examination centers around the advancement of such examinations, while featuring existing arrangements, difficulties, and perceptions shared by different exploration gatherings. Furthermore, given the restricted measure of computerized examinations performed on Romanian phony news, we review the materialness of the accessible methodologies in the Romanian setting, while at the same time recognizing future exploration ways. Alim Al Ayub Ahmed (2020) [21] web is one of the significant developments and countless people are its clients. These people utilize this for various purposes. There are diverse web-based media stages that are open to these clients. Any client can make a post or spread the word through these online stages. These stages don't confirm the clients or their posts. So a portion of the clients attempt to get out counterfeit word through these stages. This phony news can be a promulgation against an individual, society, association or ideological group. A person can't distinguish every one of these phony news. So there is a requirement for AI classifiers [22] that can recognize this phony news naturally. Utilization of AI classifiers for distinguishing the phony news is depicted in this methodical writing survey.

Razan Masood (2018) [23] fake news has created uproar recently, and this term is the Collins Dictionary Word of the Year 2017. As the news are dispersed extremely quick in the period of interpersonal organizations, a robotized reality checking device turns into a prerequisite. Notwithstanding, a completely computerized instrument that passes judgment on a case to be valid or bogus is constantly restricted in usefulness, exactness and understandability. Hence, an elective idea is to team up various investigation apparatuses in one stage which help human actuality checkers and ordinary clients produce better making a decision about dependent on numerous perspectives. A position recognition instrument is a first phase of an online test that means to identify counterfeit news. The objective is to decide the overall point of view of a news story towards its title. In this paper, we tackle the test of position identification by using customary AI calculations alongside issue explicit element designing. Our outcomes show that these models beat the best results of the taking an interest arrangements which primarily utilize profound learning models.

Sohan De Sarkar (2018) [24] satirical news identification is significant to forestall the spread of deception over the Internet. Existing ways to deal with catch news parody use AI models, for example, SVM and various leveled neural organizations alongside hand-designed highlights, yet don't investigate sentence and archive distinction. This paper proposes a strong, progressive profound neural organization approach for parody identification, which is fit for catching parody both at the sentence level and at the report level. The engineering fuses pluggable [25] nonexclusive neural organizations like CNN, GRU, and LSTM. Test results on genuine news parody dataset show significant execution gains exhibiting the adequacy of our proposed approach. An assessment of the learned models uncovers the presence of key sentences that control the presence of parody in news.

Abdullah-All-Tanvir (2019) [26] social media collaboration particularly the word getting out around the organization is an extraordinary wellspring of data these days. From one's viewpoint, its immaterial effort, direct access, and fast scattering of data that lead individuals to watch out and gobble up news from web based life. Twitter being a champion among the most notable continuous news sources also winds up a champion among the most predominant news transmitting mediums. It is known to cause broad damage by spreading pieces of tattle beforehand. Online customers [27] are typically defenseless and will, by and large, see all that they run over electronic systems administration media as solid. Therefore, automating fake news acknowledgment is rudimentary to keep up generous online media and casual association. This paper proposes a model for perceiving fashioned news messages from twitter posts, by sorting out some way to expect exactness examinations, considering automating manufactured news distinguishing proof in Twitter datasets. Subsequently, we played out an examination between five notable Machine Learning calculations, similar to Support Vector Machine, Naïve Bayes Method, Logistic Regression and Recurrent Neural Network models, independently to show the proficiency of the characterization execution on the dataset. Our exploratory outcome showed that SVM and Naïve Bayes classifier outflanks different calculations.

Hadeer Ahmed, (2017) [28] fake news is a marvel which is essentially affecting our public activity, specifically in the political world. Fake news location is an arising research region which is acquiring interest yet elaborate a few difficulties because of the restricted measure of assets (i.e., datasets, distributed writing) accessible. We propose in this paper, a fake news recognition model that utilization n-gram examination and AI [29] strategies. We examine and think about two distinct highlights extraction methods and six diverse machine arrangement strategies. Trial assessment yields the best presentation utilizing Term Frequency-Inverted Document Frequency (TF-IDF) as highlight extraction procedure, and Linear Support Vector Machine (LSVM) as a classifier, with a precision of 94%.

III. Methodology

3.1 Naive Bayes Classifier Introductory Overview

The Naive Bayes Classifier procedure depends on the supposed Bayesian hypothesis and is especially fit when the dimensionality of the information sources is high. Notwithstanding its straightforwardness, Naive Bayes can frequently beat more complex arrangement strategies. To show the idea of Naïve Bayes Classification, consider the model showed in the outline above. As demonstrated, the items can be delegated either GREEN or RED. Our assignment is [30] to arrange new cases as they show up, i.e., choose to which class name they have a place, in view of the present leaving objects shown in figure 1.

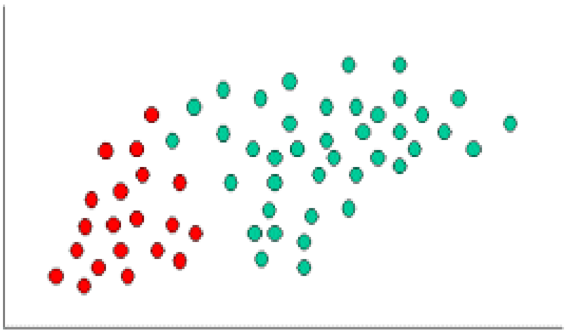

Figure1. The Naïve Bayes Classification I

Since there are twice as many GREEN items as RED, it is sensible to accept that another case (which hasn't been noticed at this point) is twice as liable to have enrollment GREEN instead of RED. In the Bayesian investigation, this conviction is known as the earlier likelihood. Earlier probabilities depend on past experience, for this situation the level of GREEN and RED items, and frequently used to anticipate results before they really occur.

Thus, we can write:

$$\text{Prior probability for GREEN} \propto \frac{\text{Number of GREEN objects}}{\text{Total number of objects}}$$

$$\text{Prior probability for RED} \propto \frac{\text{Number of RED objects}}{\text{Total number of objects}}$$

Since there is a total of 60 objects, 40 of which are GREEN and 20 RED, our prior probabilities for class membership are:

$$\text{Prior probability for GREEN} \propto \frac{40}{60}$$

$$\text{Prior probability for RED} \propto \frac{20}{60}$$

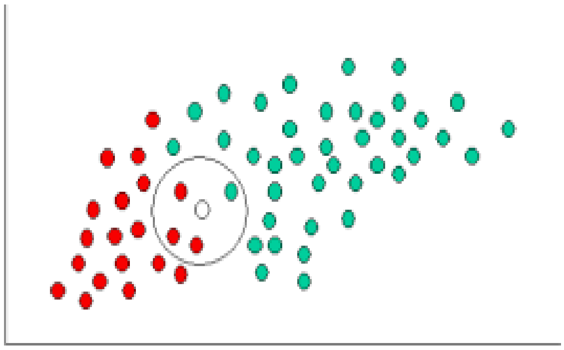

Figure 2. The Naïve Bayes Classification II

Having detailed our earlier likelihood, we are currently prepared to group [31] another item (WHITE circle). Since the articles are very much bunched, it is sensible to expect to be that the more GREEN (or RED) objects nearby X, the more probable that the new cases have a place with that specific tone shown in figure 2. To quantify this probability, we draw a circle around X which envelops a number (to be picked deduced) of focuses regardless of their group names [32]. Then, at that point we ascertain the quantity of focuses in the circle having a place with each class name. From this we figure the probability:

$$\text{Likelihood of X given GREEN} \propto \frac{\text{Number of GREEN in the vicinity of X}}{\text{Total number of GREEN cases}}$$

$$\text{Likelihood of X given RED} \propto \frac{\text{Number of RED in the vicinity of X}}{\text{Total number of RED cases}}$$

From the illustration above, it is clear that Likelihood of X given GREEN is smaller than Likelihood of X given RED, since the circle encompasses 1 GREEN object and 3 RED ones. Thus:

$$\text{Probability of } X \text{ given GREEN} \propto \frac{1}{40}$$

$$\text{Probability of } X \text{ given RED} \propto \frac{3}{20}$$

Albeit the earlier probabilities show that X may have a place with GREEN (given that there are twice as many GREEN contrasted with RED) the probability demonstrates something else; that the class participation of X is RED (given that there are more

RED articles [33] nearby X than GREEN). In the Bayesian investigation, the last characterization is created by consolidating the two wellsprings of data, i.e., the earlier and the probability, to frame a back likelihood utilizing the alleged Bayes' standard (named after Rev. Thomas Bayes 1702-1761).

$$\textit{Posterior probability of } X \textit{ being GREEN} \propto$$
$$\textit{Prior probability of GREEN} \times \textit{Likelihood of } X \textit{ given GREE}$$
$$= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60}$$
$$\textit{Posterior probability of } X \textit{ being RED} \propto$$
$$\textit{Prior probability of RED} \times \textit{Likelihood of } X \textit{ given RED}$$
$$= \frac{2}{6} \times \frac{3}{20} = \frac{1}{20}$$

Finally, we classify X as RED since its class membership achieves the largest posterior probability. The above probabilities are not normalized. However, this does not affect the classification outcome since their normalizing constants are the same.

3.2 Technical Notes

In the previous section, we provided an easy-to-understand example of classification using Naive Bayes. This section contains additional information about the technical issues at hand. Naive Bayes classifiers can handle an arbitrary number of independent variables [34] whether continuous or categorical. Given a set of variables, X = {x1,x2,x...,xd}, we want to construct the posterior probability for the event Cj among a set of possible outcomes C = {c1,c2,c...,cd}. In a more familiar language, X is the predictors and C is the set of categorical levels present in the dependent variable. Using Bayes' rule:

$$p(C_j \mid x_1, x_2, \ldots, x_d) \propto p(x_1, x_2, \ldots, x_d \mid C_j) p(C_j)$$

where p(Cj | x1,x2,x...,xd) is the posterior probability of class membership, i.e., the probability that X belongs to Cj. Since Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent we can decompose the likelihood to a product of terms:

$$p(X \mid C_j) \propto \prod_{k=1}^{d} p(x_k \mid C_j)$$

and rewrite the posterior as:

$$p(C_j \mid X) \propto p(C_j) \prod_{k=1}^{d} p(x_k \mid C_j)$$

Utilizing Bayes' standard above, we name another case X with a class level Cj that accomplishes the most noteworthy back likelihood. Albeit the suspicion that the indicator (free) factors are autonomous isn't generally exact, it improves on the grouping task drastically, since it permits the class contingent densities p(xk | Cj) to be determined independently for every factor, i.e., it diminishes a multidimensional errand to various one-dimensional ones [35]. As a result, Naive Bayes diminishes a high-dimensional thickness assessment undertaking to a one-dimensional piece thickness assessment. Moreover, the suspicion doesn't appear to incredibly influence the back probabilities, particularly in areas close to choice limits, in this manner, leaving the arrangement task unaffected. Innocent Bayes can be demonstrated in a few distinct manners including ordinary, lognormal, gamma and Poisson thickness.

3.4 Decision Tree Algorithm

Decision Tree algorithm has a place with the group of managed learning calculations. In contrast to other administered learning calculations, the decision tree calculation can be utilized for tackling relapse and order issues as well. The objective of utilizing a Decision Tree is to make a preparation model that can use to anticipate the class or worth of the objective variable by taking in basic decision principles surmised from earlier data [36] (training information). In Decision Trees, for anticipating a class name for a record we start from the foundation of the tree. We think about the upsides of the root [37] trait with the record's characteristic. Based on examination, we follow the branch relating to that worth and leap to the following hub.

IV. CONCLUSION

So, was your fake news classifier investigation a success? Definitely not. Regardless, will you experiment with another dataset, try out some NLP order models, and assess how fruitful they were? Indeed. Characterizing fake news with simple pack of words or TF-IDF vectors is a distorted methodology that has been in place since the beginning. Especially with a multilingual dataset full of boisterous tokens.

If you hadn't investigated what your model had truly learned, you might have assumed the model had mastered something significant. As a result, remember to consistently examine your models. I'd be interested if you discovered any patterns in the data that I may have overlooked. As far as significant highlights on my blog, I'll be returning to a post on how various classifiers analyse. If you spend some time exploring and discover something interesting, please share your findings and notes in the comments section, or you can generally connect on dataset. Our preliminary plan is to investigate a larger dataset to see how the traditional model, such as Naive Bayes, competes with profoundly computational neural organisation based models to detect fake news.

V. REFERENCES

[1]. Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pages 127–138. Springer, 2017.

[2]. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2):211–36, 2017.

[3]. Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, pages 15–19. ACM, 2015.

[4]. Yusuf Perwej, Firoj Parwej, Mumdouh Mirghani Mohamed Hassan, Nikhat Akhtar, "The Internet-of-Things (IoT) Security: A Technological Perspective and Review" , International Journal of Scientific Research in Computer Science Engineering and Information Technology, Volume 5, Issue 1, Pages 462-482, 2019, DOI: 10.32628/CSEIT195193

[5]. Mathieu Cliche. The sarcasm detector, 2014

[6]. Johannes F¨urnkranz. A study using n-gram features for text categorization. Austrian Research Institute for Artifical Intelligence, 3(1998):1–10, 1998

[7]. Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 84–89, 2017

[8]. Song Feng, Ritwik Banerjee, and Yejin Choi. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, pages 171–175. Association for Computational Linguistics, 2012.

[9]. Al-Mushayt, O.S., Perwej, Y. and Haq, K.," Electronic-government in Saudi Arabia: A positive revolution in the Peninsula", arXiv preprint arXiv:1205.3986, 2012

[10]. Yusuf Perwej, "The Bidirectional Long-Short-Term Memory Neural Network based Word Retrieval for Arabic Documents", Transactions on Machine Learning and Artificial Intelligence

(TMLAI), Volume 3, Issue 1, Pages 16 - 27, 2015, DOI: 10.14738/tmlai.31.86

[11]. A ngel Herna´ndez-Castan˜eda and HiramCalvo. Deceptive text detection using continuous semantic spacemodels. Intelligent Data Analysis, 21(3):679–695, 2017

[12]. Yusuf Perwej , Firoj Parwej, "A Neuroplasticity (Brain Plasticity) Approach to Use in Artificial Neural Network", International Journal of Scientific & Engineering Research (IJSER), France , ISSN 2229 – 5518, Volume 3, Issue 6, Pages 1- 9, 2012, DOI: 10.13140/2.1.1693.2808

[13]. Nikhat Akhtar, Devendera Agarwal, "An Efficient Mining for Recommendation System for Academics", International Journal of Recent Technology and Engineering (IJRTE), SCOPUS, Volume-8, Issue-5, Pages 1619-1626, 2020 , DOI: 10.35940/ijrte.E5924.018520

[14]. Ethar Qawasmeh, Mais Tawalbeh, Malak Abdullah, "Automatic Identification of Fake News Using Deep Learning", 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 978-1-7281-2946-4/19/$31.00 ©2019 IEEE

[15]. William Yang Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection", arXiv:1705.00648v1 cs.CL1 May 2017

[16]. Yusuf Perwej, Shaikh Abdul Hannan, Nikhat Akhtar, "The State-of-the-Art Handwritten Recognition of Arabic Script Using Simplified Fuzzy ARTMAP and Hidden Markov Models", International Journal of Computer Science and Telecommunications, Volume, Issue 8, Pages 26- 32, 2014

[17]. Perwej A, Yadav KP, Sood V, Perwej Y,"An evolutionary approach to bombay stock exchange prediction with deep learning technique", IOSR J Bus Manag (IOSR-JBM) 20(12):63–79, 2018

[18]. Z Khanam1 , B N Alwasel1 , H Sirafi1 and M Rashid, "Fake News Detection Using Machine Learning Approaches", IOP Conf. Series: Materials Science and Engineering 1099 (2021) 012040 IOP Publishing doi:10.1088/1757-899X/1099/1/012040.

[19]. Y. Perwej, Ashish Chaturvedi, "Machine Recognition of Hand Written Characters using Neural Networks", International Journal of Computer Applications (IJCA), USA, ISSN 0975 – 8887, Volume 14, No. 2, Pages 6- 9, 2011, DOI: 10.5120/1819-2380

[20]. Costin BUSIOC, Stefan RUSETI, Mihai DASCALU, " A Literature Review of NLP Approaches to Fake News Detection and Their Applicability to Romanian- Language News Analysis", 2020, Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-1794, within PNCDI III.

[21]. Alim Al Ayub Ahmed, A. Aljarbouh, Praveen Kumar Donepudi," Detecting Fake News using Machine Learning: A Systematic Literature Review", IEEE conference, 2020

[22]. Perwej, Y., Perwej, A.: Prediction of the Bombay Stock Exchange (BSE) market returns using artificial neural network and genetic algorithm. J. Intell. Learn. Syst. Appl. 4, 108–119, 2012

[23]. Razan Masood and Ahmet Aker," The Fake News Challenge: Stance Detection using Traditional Machine Learning Approaches", In Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KMIS 2018), pages 128-135

[24]. Sohan De Sarkar, Fan Yang," Attending Sentences to detect Satirical Fake News", Proceedings of the 27th International Conference on Computational Linguistics, pages 3371–3380 Santa Fe, New Mexico, USA, August 20-26, 2018.

[25]. Asif Perwej, Y. Perwej, Nikhat Akhtar, and Firoj Parwej, "A FLANN and RBF with PSO

Viewpoint to Identify a Model for Competent Forecasting Bombay Stock Exchange", COMPUSOFT, SCOPUS, An International Journal of Advanced Computer Technology, 4 (1), Volume-IV, Issue-I, Pages 1454-1461, 2015, DOI: 10.6084/ijact.v4i1.60
[26]. Abdullah-All-Tanvir, Ehesas Mia Mahir, Saima Akhter" Detecting Fake News using Machine Learning and Deep Learning Algorithms", 2019 7th International Conference on Smart Computing & Communications (ICSCC).
[27]. Poranki KR, Perwej Y, Perwej A. ," The Level of Customer Satisfaction related to GSM in India", RJSITM., 4(3):32-3, 2015
[28]. Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, page 82. American Society for Information Science, 2015.
[29]. Yusuf Perwej, Nikhat Akhtar, Firoj Parwej,"The Kingdom of Saudi Arabia Vehicle License Plate Recognition using Learning Vector Quantization Artificial Neural Network", International Journal of Computer Applications (IJCA), USA, ISSN 0975 – 8887, Volume 98, No.11, Pages 32 – 38, July 2014, DOI: 10.5120/17230-7556
[30]. Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 4647–4657. ACM, 2016.
[31]. Shlok Gilda. Evaluating machine learning algorithms for fake news detection. In Research and Development (SCOReD), 2017 IEEE 15th Student Conference on, pages 110–115. IEEE, 2017.
[32]. Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. In Electrical and Computer Engineering (UKRCON), 2017 IEEE First Ukraine Conference on, pages 900–903. IEEE, 2017.
[33]. Johan Hovold. Naive bayes spam filtering using word-position-based attributes. In CEAS, pages 41–48, 2005.
[34]. Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759, 2016.
[35]. Hadeer Ahmed, Issa Traore, Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques",International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments ISDDC 2017: Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments pp 127-138
[36]. Y. Perwej, N. Akhtar, Firoj Parwej,"The Kingdom of Saudi Arabia Vehicle License Plate Recognition using Learning Vector Quantization Artificial Neural Network", International Journal of Computer Applications (IJCA), USA, ISSN 0975 – 8887, Volume 98, No.11, Pages 32 – 38, 2014, DOI: 10.5120/17230-7556
[37]. L. Rokach and M. Oded, Decision Trees: Theory and Applications, Singapore:World Scientific, 2008

Cite this article as :

Vaishali Singh, "A Review of Machine Learning-Based Fake News Analysis", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 2, pp. 252-261, March-April 2022. Available at
doi : https://doi.org/10.32628/IJSRST229245
Journal URL : https://ijsrst.com/IJSRST229245