International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5921
Article Received: 10 October 2022 Revised: 14 November 2022 Accepted: 22 December 2022

Applications of Machine Learning for Fake News Detection in Social Networks

[1]Dr. Ch. N. Santhosh Kumar, [2]Dr. M. Sailaja, [3]Dr. Md. Ali Hussain, [4]Dr. Syed Ziaur Rahman
[1]Associate Professor, Dept. Of Computer Science and Engineering
GITAM (Deemed to be University), Hyderabad, India
santhosh.ph10@gmail.com
[2]Assistant Professor , Dept of CSE
Prasad V Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada
msailaja@pvpsiddhartha.ac.in
[3]Professor, Dept. of ECM, KoneruLakshmaiahEducationFoundation, Vadeswaram, Gnt Dt,
Andhra Pradesh, dralihussain@kluniversity.in
[4]Assistant Professor, Faculty of Information Technology, Majan University College, Muscat,
Sultanate of Oman.
syed.rahman@majancollege.edu.om

Abstract: The value of online media for getting news is questionable. People seek out and devour news from online media because it is convenient, inexpensive, and widely disseminated. In contrast, it facilitates the widespread distribution of "counterfeit news," or news of lower quality that includes fabricated data. Many people and institutions are negatively impacted by the widespread circulation of false information. As a result, detecting fake news via social media has emerged as a topic of interest for academics. Searching for and reading the news is becoming increasingly convenient as a result of the widespread availability, quick expansion, and widespread dissemination of traditional news outlets and social media. Nowadays, there is a plethora of information that can be found on social media, and it can be difficult to tell what is real and what is not. The distribution costs of releasing news via social media are inexpensive, and anyone can do it. The widespread circulation of false information could have devastating effects on both individuals and communities. Developing a reliable machine learning method for spotting fake news is the focus of this work.

Keywords: Fake news, Sentimental Analysis, News, Social Network.

I. INTRODUCTION

Misleading information that can be verified is often found in fake news. This perpetuates a lie about a certain statistic in a country or the inflated price of a specific service for a country, both of which can spark unrest, as seen in the Arab Spring. The House of Commons and the Crosscheck project are only two examples of institutions working to address problems such as verifying the author's credibility. However, their applicability is restricted since they rely on human manual detection, which is neither accountable nor viable in a world where millions of items are either purged or published every minute. One potential answer is the creation of an automated index or rating system that assesses the veracity of various news sources and publishers.

[1,2] A reduction in the reinforcement learning needed to manage sentimental data analysis for social networks based on human text data and emotional data is discussed, along with the prospect of learning opportunities that might aid. This paves the way for an intelligent approach that effectively determines if a news text is phoney or not. In recent days, social Twitter have evaluated and discussed how vast volumes of data are being considered and processed, allowing for successful identification of fresh insights. These considerations contribute to the framework's efficient deep learning. Taking into account the issue of heterogeneity, we propose the effective Sentimental based Decision Making Approach (SDMA) to identify the fake news review from the social network using Deep Learning. It aids in analysing the dataset's news reviews to determine the prevailing sentiment based on the reviewers' expressed feelings and words. Given the heterogeneity issue in data processing, the data can be extracted according to a predetermined set of criteria, resulting in a new data representation. To demonstrate how the effective prediction helps to identify the trained data classification via sentimental analysis by deploying the SDMA approach for social networks, we will use social information to monitor the prediction phase through a modified auto-encoder i.e., Deep Learning.

Data extraction is aided by pre-processing, which involves eliminating special characters from the text, filtering data to eliminate duplication, filtering emotional text, and erasing URLs. Then, it aids in enhancing the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5921
***Article Received**: 10 October 2022 **Revised**: 14 November 2022 **Accepted**: 22 December 2022*

chosen operations during the training phase with a learning algorithm to foresee the process in advance, which aids in identifying and mitigating the treat, which affects human interaction just before they are about to share the information. This proposed study modifies the auto-encoder algorithm by adding new parameters for better text classification and tweet filtering. The learning procedure facilitates the collection and generation of voluminous data in a continuous stream. Taking into account news/tweets in the social network aids in improving the method applied for the time-consuming task of determining different metrics of human intervention. Data classification using sentiment analysis helps to improve secure text messaging based on monitoring and resource utilisation.

This work provides a way for developing a model that can determine whether an article is real or not by analysing its words, phrases, sources, and titles, with the help of supervised machine learning algorithms applied to a manually sorted and guaranteed dataset. The outputs of the confusion matrix are then used as input to feature selection algorithms, which are then used to experimentally pick the most optimal features in order to get the highest accuracy. Our plan is to build the model by combining various existing classification methods. As a result of the model's analysis of the unseen data and the subsequent visualisation of the results, the final outcome will be a model for the identification and classification of bogus articles that can be implemented in any system.

II. RELATED WORKS

Recently, sentiment analysis and learning methodologies have been used to great effect in the improvement of social networks [5]. In this instance, the social network procedure is included with multiple tweets and SMS messages conveying various sentiments. Learning technologies based on real-time procedures like data filtering and erasure are the subject of a large-scale survey [6] and [7]. The social datasets are taken into account as input datasets, which aids in the provision of intelligence to collect the data, i.e., data extraction, data processing, and training those data. Tweet/text fake identification based on heterogeneity is discussed [8] with a focus on finding a workable solution for social media. This point is crucial for the analysis [9] Using a constant ergodic Markov chain as a training sample, Support Vector Machine (SVM) classification learning techniques are robust. With the help of Markov samples and the knowledge that the kernel Hilbert spaces tend to recur in certain patterns, they were able to construct an acceptable bound on the misclassification error of an online SVM ordering algorithm. This lends credence to the proposed enhancements to the learning process.

Since social media plays such a major role, [10] discusses how to keep an eye out for fake news on these platforms and how to manage the content posted there. It will point the way toward developing the strategy, which will incorporate machine learning and sentimental data classification. In order to determine whether or whether information shared on social media platforms poses a security risk, it is necessary to use a method based on sentimental analysis [11] in conjunction with learning techniques. This paves the way for the possibility of combining the data extraction and pre-processing stages of the learning approach.

[12] Talk about how the suggested technique uses a neural network-based learning algorithm to take into account the tweets and messages that are widely shared on social media. The integration of Bayesian learning with neural networks aids in raising the accuracy of predictions. [13] The Internet of Things (IoT) and health care have been proposed as a means of automating everyday activity monitoring among the elderly via various sensor devices. In this case, the framework incorporates user preferences into the data mining process in order to keep tabs on the patient's routine activity [14]. Multiple learning algorithms, including GRU-SVM, Linear Regression, Multilayer Perceptron (MLP), Nearest Neighbor (NN) search, Softmax Regression, and Support Vector Machine (SVM) [16], use data from the learning repository dataset [15] to compute accuracy, test datasets, and sensitivity and specificity.

Characteristics are included in the datasets that make up the training and validating data sets that are referenced in [17]'s learning repository. These attributes include input and output variables. Textual information, feelings, online resources (URLs), repetition, etc., all contribute to the final product. To improve data accuracy, specificity, etc., existing research works typically employ Sentiment-Analysis-on-Twitter-Dataset-using-Learning-datasets. In order to extract the knowledge data by applying the back propagation neural networks for social applications, Sentiment-Analysis-on-Twitter-Dataset-using-Learning-datasets is considered [18]. Then, K-means and support vector machine (SVM) are combined in [19] for feature extraction, which is then applied to the social networks to gauge their reliability. Data mining algorithms like LR, Neural networks, and Decision trees are compared in order to find the most accurate predictor of fake social media posts.

Some data's can be used to run a more in-depth detection process in the future [20]. Biometric data, facial

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5921
***Article Received**: 10 October 2022 **Revised**: 14 November 2022 **Accepted**: 22 December 2022*

expression, and speech-based ECG signals based on social data can all be taken into account and analysed to facilitate an efficient detection procedure. It is possible to tweak other learning algorithms with different social data in order to test the data's precision and accuracy in different contexts. Therefore, taking into account the issue of heterogeneity, we propose the effective Sentimental based Decision Making Approach (SDMA) to identify the fake news / review from the social network via learning technique. It's useful for analysing the dataset's reviews and making predictions about the prevailing sentiment based on the feelings and language used by the news sources and reviewers.

III. RESEARCH METHODOLOGY

The following figure 1, sourced from google.com, details the machine learning techniques we hope to employ in our work to determine whether or not a given piece of news is fake. There was some preliminary processing of raw news texts that was necessary after we gathered the necessary data. for this reason, we need to pre-process the data. There is a strong correlation between how well-designed features are and how well machine learning models perform. In order to train the data for classification, we must first extract a large variety of characteristics. The next step is to determine whether or not a set of data is fake.

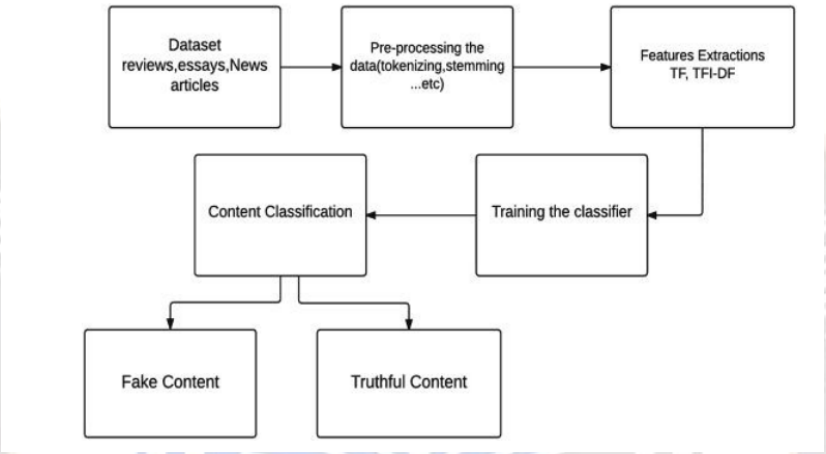

Fig 1: Typical Framework for fake news detection using machine leaning techniques

The Naive Bayes, Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine models on the Liar Dataset are currently being used in our work.

The Naive Bayes approach is a family of supervised learning algorithms that uses Bayes theorem as its theoretical foundation. The probability of an occurrence is calculated using Bayes' theorem when we know the probability of a prior event. Naive Bayes relies heavily on the notion that each characteristic can be evaluated separately.

The random forest, also known as the random decision forest, is a supervised machine learning algorithm for classification and regression that uses ensemble learning to build many decision trees at once during training and then use those trees to determine the most common classes. Better results can be obtained and overfitting is avoided if there are more trees in the data set. The goal is to improve prediction accuracy by combining these results. Each decision tree receives its own data-driven training. A single more data point will not alter the algorithm's performance.

In Machine Learning, the likelihood of a categorical dependent variable can be predicted using a classification process called Logistic Regression. It is important to note that the dependent variable in logistic regression is a binary one, meaning that it can only take on the values 1 (yes, a successful outcome) or 0.

If you're having trouble making up your mind, a decision tree can help you visualise the various outcomes that could arise from your choices, as well as the associated costs and benefits. This is one representation for an algorithm consisting entirely of if/then statements.

IV. RESULTSAND DISCUSSIONS

To that end, we'll be using the political news data from the Liar-dataset, which has been annotated as either fake or trust news and is a New Benchmark Dataset for Fake News Detection, to conduct our analysis. The "Liar" dataset has been analysed by us. The confusion matrix shows the outcomes of the dataset analyses performed by the six algorithms. These six detection algorithms are as follows:

- XGboost.
- Random Forests.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5921
***Article Received**: 10 October 2022 **Revised**: 14 November 2022 **Accepted**: 22 December 2022*

- Naive Bayes.
- K-Nearest Neighbors (KNN).
- Decision Tree.
- SVM

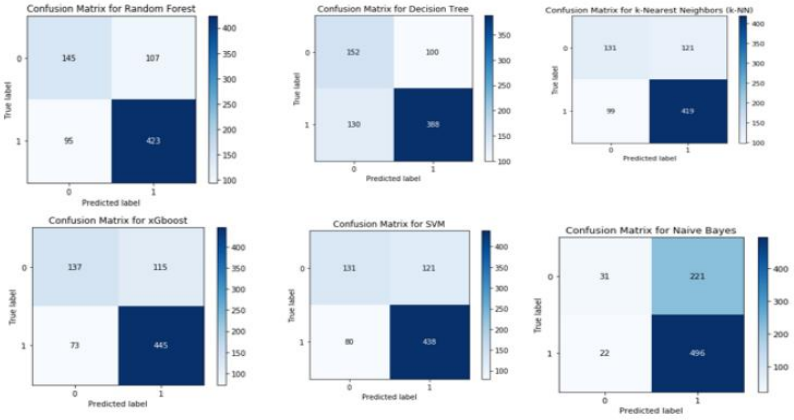

Figure 2. Confusion Matrix Results for the Algorithms

All of the algorithms' Confusion Matrixes are shown in Figure 2.

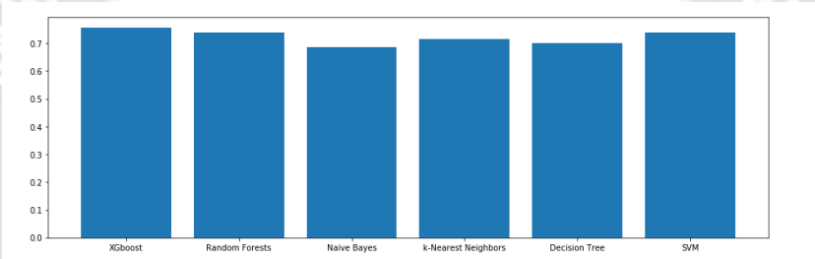

Figure 3. Accuracy Results of all the Algorithms

The precision of these algorithms is depicted in Figure 3. XGBOOST depicts the highest accuracy, at over 75%, followed by SVM and Random forest, each with around 73% accuracy.

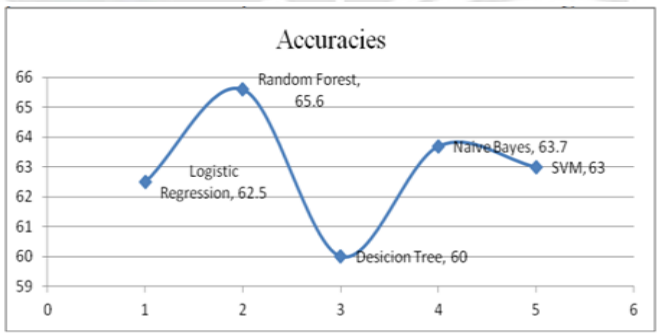

Fig 4: Typical Framework for fake news detection using machine leaning techniques.

Accuracy is plotted along the y axis of Figure 4, while the x axis shows the many machine learning models tried.

V. CONCLUSION

Most people today get their news from social media sites like Facebook and Twitter rather than newspapers and television. People's lives and communities suffer greatly as a result of this spread of false information. As a result, we conduct a literature review aimed at determining the most effective methods for detecting fake news, and conclude that Word Embedding, Tokenization, and Parts of Speech tagging are the most effective pre-processing techniques, while TF-IDF and Count Vectorizer provide the most useful features. As a result, we plan to employ these strategies for PreProcessing and feature extraction as part of a more robust strategy, as well as the Random Forest classifier, Convolutional Neural Networks, Long Short Term Memory, and an Ensemble Learning Approach to improve accuracy.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5921
***Article Received**: 10 October 2022 **Revised**: 14 November 2022 **Accepted**: 22 December 2022*

REFERENCES

[1]. Regin, R. and Rajest, S. Suman and Singh, Bhopendra (2021) Fault Detection in Wireless Sensor Network Based on Deep Learning Algorithms. EAI Transactions on Scalable Information Systems. e8. ISSN 2032-9407

[2]. Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou and Ping Li, User-level sentiment analysis incorporating social networks, KDD '11: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp.1397–1405, https://doi.org/10.1145/2020408.2020614

[3]. Kumar, C., Bharati, T.S. & Prakash, S. Online Social Network Security: A Comparative Review Using Machine Learning and Deep Learning. Neural Process Lett 53, 843–861 (2021). https://doi.org/10.1007/s11063-020-10416-3.

[4]. Akinyelu, Andronicus A. Advances in spam detection for email spam, web spam, social network spam, and review spam: ML-based and nature-inspired-based techniques, Journal of Computer Security, vol. 29, no. 5, pp. 473-529, 2021.

[5]. Yang Liu, Lei Ma & Jianjun Zhao, Secure Deep Learning Engineering: A Road Towards Quality Assurance of Intelligent Systems, Lecture Notes in Computer Science, pp.3-15, 2019.

[6]. Xinyan Li; Yufei Chen; Cong Wang; Chao Shen, When Deep Learning Meets Differential Privacy: Privacy,Security, and More, IEEE Network, Vol. 35, Issue. 6, pp. 148-155, 2021.

[7]. Yu Zhang and Ershi Qi, Happy work: Improving enterprise human resource management by predicting workers' stress using deep learning, PLOS ONE, 2022.

[8]. Haoxiang Wang; Guihuang Liang; Xingming Zhang, Feature Regularization and Deep Learning for Human Resource Recommendation, IEEE Access, Vol. 6, pp. 39415-39421, 2018.

[9]. Hyeoncheol Lee, Youngsub Han, Kwangmi Ko Kim, Sentiment Analysis on Online Social Network Using Probability Model, AFIN 2014: The Sixth International Conference on Advances in Future Internet, 2014.

[10]. JingJing Zenga, Yuze Duan, Desheng Wang, Bin Zou, Yue Yin and Jie Xu, Generalization performance of Lagrangian support vector machine based on Markov sampling, Journal of Statistical Planning and Inference, Vol. 214, pp. 89-104, 2021.

[11]. Guang Yang; Shibo He; Zhiguo Shi, Leveraging Crowdsourcing for Efficient Malicious Users Detection in Large-Scale Social Networks, IEEE Internet of Things Journal, Vol, 4, Issue: 2, pp. 330-339, April 2017.

[12]. Nayomi Kankanamge, Tan Yigitcanlar, Ashantha Goonetilleke and Md.Kamruzzaman, Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets, International Journal of Disaster Risk Reduction, Vol. 42, 2020.

[13]. Francesco Pierri, Stefano Ceri, False News On Social Media: A Data-Driven Survey Share on, ACM SIGMOD RecordVolume 48Issue 2June 2019 pp 18–27.

[14]. Abdulhamit Subasi; Mariam Radhwan; Rabea Kurdi; Kholoud Khateeb, IoT based mobile healthcare system for human activity recognition, 2018 15th Learning and Technology Conference (L&T), 2018.

[15]. Farman Alia, ShakerEl-Sappagh, S.M. Riazul Islam, Amjad Ali, Muhammad Attique, Muhammad Imran, Kyung-Sup Kwak, An intelligent healthcare monitoring framework using wearable sensors and social networking data, Future Generation Computer Systems, Vol. 114, pp. 23-43, 2021.

[16]. Jiban Khuntia, Hang Sun & Dobin Yim, Sharing News through Social Networks, International Journal on Media Management, Vol.8, Issue. 1, pp.59 – 74, 2016.

[17]. Mohammad Mahmudur Rahman Khan; Rezoana Bente Arif; Md. Abu Bakr Siddique; Mahjabin Rahman Oishe, Study and Observation of the Variation of Accuracies of KNN, SVM, LMNN, ENN Algorithms on Eleven Different Datasets from UCI Machine Learning Repository, 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), 2018.

[18]. Tinglong Tang, Shengyong Chen, Meng Zhao, Wei Huang & Jake Luo, Very large-scale data classification based on K-means clustering and multi-kernel SVM, Soft Computing, Vol. 23, pp. 3793–3801, 2019.

[19]. Konstantinas Korovkinas, Paulius Danenas And Gintautas Garsva, SVM and k-Means Hybrid Method for Textual Data Sentiment Analysis, Baltic J. Modern Computing, Vol. 7, 2019, No. 1, pp. 47–60.

[20]. Zhang Rui; Zheng Yan, A Survey on Biometric Authentication: Toward Secure and Privacy-Preserving Identification, IEEE Access, Vol. 7, pp. 5994 – 6009, 2018.