Machine Learned Classifiers for Trustworthiness Assessment of Web Information Contents

Priyanka Meel
Biometric Research Laboratory
Department of Information technology
Delhi Technological University
New Delhi, India-110042
priyankameel86@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory
Department of Information technology
Delhi Technological University
New Delhi, India-110042
dvishwakarma@gmail.com

2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) | 978-1-7281-8529-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCIS51004.2021.9397228

***Abstract*—Social networking, information sharing, knowledge imparting, discussions on current happenings etc. are always a part of human society. With the fast pace of life and advancement in technology; people rely more on online information, as a result of this web platforms have become a dominant place for social interactions. This has given rise to unverified and unauthenticated news that has extremely negative effects. Fake news, rumor, misinformation, disinformation, satire, hoax, clickbait, propaganda are all different flavors of the same malice of information pollution. The research community is constantly trying to figure out a viable technical solution to this problem in different ways. In this work, we designed a framework based on five independent supervised machine-learned classifiers Support Vector Machine, K-Nearest Neighbor, Logistic Regression, Naïve Bayes and Random Forest for trustworthiness assessment of web information contents. The classifiers are being trained and tested on two different datasets: Fake News Detection (Jruvika/FND) and Real or Fake News that contains full news articles in the form of headline and body. Experiments and result analysis verify that the highest accuracy attained by the projected method is 96.61% on the Fake News Detection dataset using the SVM classifier. The work is also compared with other contemporary techniques.**

Keywords— Machine Learning, Infodemic, Supervised Model, Support Vector Machine

I. INTRODUCTION

Humans have always interacted with their kind to survive, gain knowledge and pass on information about the happenings around them. So social networking existed right from the beginning of human history. With the advancement in technology and the ease of getting news and information handy and on the go, one may have to compromise with the authenticity of the information. A large section of society is using social networking sites such as Twitter, Facebook, Instagram, WhatsApp, YouTube, etc. to obtain news rather than using authenticated news channels and sources [1]. At the same time, unauthenticated sources through these social media platforms, websites and Twitter handle distribute fabricated news to the audience. In such a scenario it becomes difficult to understand on which source one should rely on to obtain unfabricated and genuine news. The misinformation or rumors spread across especially during emergencies can have a devastating effect on individuals and society. Spurious news in such a scenario would not only give rise to panic among the individuals but in some cases, it may also target a particular community.

Important events that leave a mark on the society such as elections, war, stock prices, business deals, politics, the health status of celebrities, cryptocurrency, religious events etc. noticeably give birth to fake news. The 2016 general elections in the USA were driven by fake news. After demonetization in India, the new 2000-rupee notes were falsely advertised to have a chip installed in it. The current coronavirus pandemic has given vertical rise to a surge of fake news being named as "**infodemic**" by WHO officials [2]. The key reason behind this is that dependable news sources are recurrently swallowed up by unconfirmed online information. All these events prove that fake news is a huge threat to our society and a problem that should be given utmost attention.

The younger generation is thought to be more tech-savvy than their parents, but they too seem to lack the ability to tell whether a piece of particular news is fabricated or genuine. The research done by Common Sense Media puts forward the fact that 44% of them have confirmed that they cannot differentiate between fake and genuine news. While it was also found that at least one online news story is being shared every day by 31% of kids aged between 10 to 18 that they found out later was inaccurate or fake [3].

The intention of sharing spurious news can vary. Few people share fake news to gain profit, while others share it to defame the opponent, some target a particular community and some try to mislead the readers and so on. During the election season, the primary motive is to mislead readers and to defame the opponents. But there are groups of people who might be sharing it to gain some monetary benefits. This was observed during the 2016 US presidential election that teenagers in the Macedonian town of Veles had gained at least $60,000 in 6 months by spreading fake news [4], [5]. So, the motive would largely vary from person to person. Researchers working in different fields are trying to provide a viable solution to this problem of misinformation and fake news that will eventually help in enhancing the reliability of online information. In this work, we tried to design an efficient framework using machine-learned classifiers to check the credibility status of online news articles. The main highlights of the work are as follows:

ISBN: 978-1-7281-8529-3/21/$31.00 ©2021 IEEE

• A framework is designed for the trustworthiness assessment of web information contents using machine-learned classifiers.

• Five machine-learned classifiers Support Vector Machine, Logistic Regression, Random Forest, Naïve Bayes and K-Nearest Neighbor are being independently trained and tested.

• Two standard Fake News datasets Fake News Detection (Jruvika/FND) and Real or Fake News are being used for training and testing purposes in the proportion of 80% and 20% respectively.

• Accuracy, Recall (Hit Rate/Sensitivity), Precision (Positive Predictive Value), F1-score, Misclassification rate (Error rate), Type-I error (False Positive Rate/fall out) and Type-II Error (False Negative Rate/Miss Rate) parameters are being used for performance assessment of classifiers.

• The performance comparison of machine learning methods is being done among themselves and also with other state-of-the-arts.

The work is organized systematically in five different sections. Section 1 focuses on the problem of fake news and its different aspects. The literature survey is presented in section 2, section 3 emphasis on the proposed framework; experimental settings, result analysis and state-of-the-art comparison are being pronounced in section 4. Lastly, section 5 accomplishes the work by highlighting possible areas of future research.

II. Related Works

Information that is being shared on the internet is in different data presentations of script, picture, audio and video. Artificial Intelligence Techniques are being constantly used by researchers to figure out different forgery formats of text and multimodal data. Rashlin et al. [6] analyzed the linguistic characteristics and writing style of fake news, hoax, satire and propaganda to differentiate it with the real articles in news media. The authors specifically focused on political fact-checking and described how the same event can be described by different platforms to gain their vested benefits out of it. Text-based word vector representations of Bag of Words, TF-IDF, Word2Vec along with stylometric text features are explored by Reddy et al. [7] using bagging and boosting ensemble methods to classify a new article as real or fake.

Sui et al. [8] presented an elaborative review on different aspects of misinformation including stance detection, fact-checking, abstractive summarization, rumor detection and emotion analysis. They also categorized the publicly available benchmark datasets such as LIAR, FEVER, BuzzFeedNews, BuzzFace, PHEME, RumourEval, CREDBANK, BS Detector, FakeNewsNet, Fake or Real for misinformation detection. A deep neural network is designed by Liu and Wu [9] tested on Twitter15 and Weibo16 dataset for fake early detection of fake news with an accuracy of 90% before it is retweeted 50 times. The architecture is based on three innovative components: status-sensitive crowd response feature extractor, position-aware attention mechanism and multi-region mean pooling mechanism. Gangireddy et al. [10] devised an unsupervised framework driven by a graph-based technique using inter-user behavior dynamics to classify fake and legitimate news. Their proposed architecture GTUT has experimented on PolitiFact and GossipCop dataset.

Abonizio et al. [11] extracted stylometric and psychological text features of news articles written in Spanish, American English and Brazilian Portuguese to design a language-independent framework to categorize legitimate, fake and satirical news. Four machine learning algorithms SVM, KNN, Random Forest and Extreme Gradient Boosting (XGB) are implemented in this method with 85.3% highest accuracy. A rumor that is termed as unverified information which may or may not be true is also a prominent area of research. Several aspects of rumor generation, propagation, detection, resolution and its real-life effects are being explored in various studies [12] [13] [14].

In line with the work done by other researchers to overcome the problem of fake news, we tried to design a simple and efficient framework using machine-learned classifiers by extracting explicit features from news text.

III. Proposed Model

Machine Learned classifiers are supervised methods in which fully labeled data is being used for training the classifiers using the explicit features extracted from training samples. This work deals with only the headline and body of the news article and tries to distinguish between the fake and real news based on features extracted from these two fields. Five supervised machine-learned classifiers are being independently trained and tested for two different datasets according to the architecture described in Figure 1 and Algorithm 1.

Algorithm 1: Working steps in Machine Learning framework

1: Import required python libraries
2: Read the data file
3: Perform data cleaning
4: Convert text into vector form using Count Vectorizer
5: Split the dataset into train and test samples
6: Perform tf-idf feature extraction
7: Train ML classifiers (SVM, LR, RF, NB, KNN) on extracted features
8: Test the trained models using test samples

A. Preprocessing, Vectorization and Feature Extraction

The preliminary steps of preparing the data for training the classifiers are pre-processing, vectorization and feature extraction. The datasets contained a lot of extra metadata so only the required columns of headline/title, text/body and label/type are extracted. Then the complete corpus is changed into lowercase; white spaces, punctuations and stop words are removed, stemming is applied and finally, the headline and body of the article are concatenated together to make it a single instance of the corpus. To convert the text into a vector form count vectorizer is used.

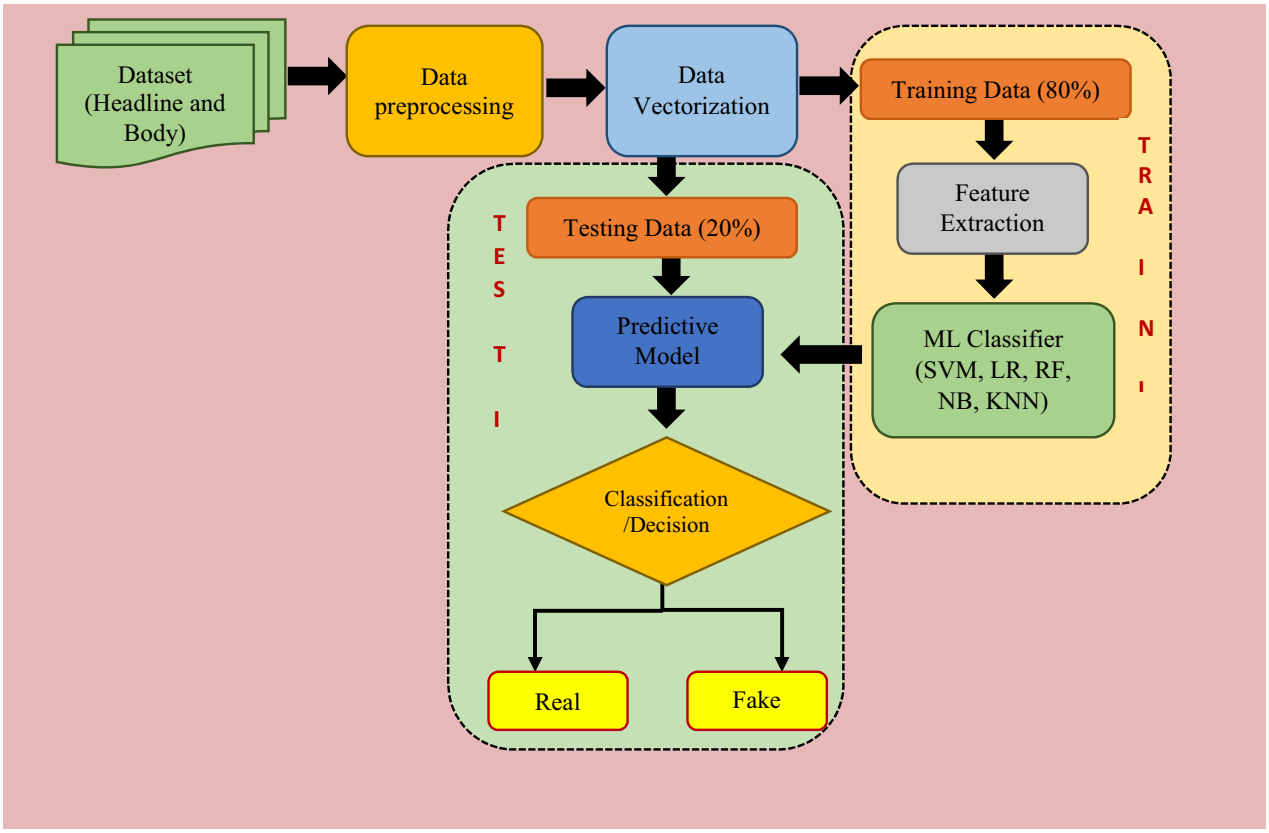

Fig. 1. Proposed Machine Learning Architecture

Feature extraction of text instances is done using tf-idf (Term Frequency-Inverse Document Frequency) which echoes the importance of each word for the given document.

B. Support Vector Machine (SVM)

Support Vector Machine follows a non-probabilistic approach as it assigns the incoming (reading) data points to a class while training, and does the same operation based on attributes learned during training in the testing phase. It is a linear classifier but can also do non-linear classification using kernel tricks (nonlinear kernel). It gives a superb performance in applications where textual data is to be classified, which makes it a really good algorithm to look at when classifying fake news. SVM model maps the training instances as points in space and classifies them in separate categories by a clear breach that is as wide as possible. Testing samples are then plotted into the identical plane and predicted to fall into a category based on the side of the breach in which they fall as illustrated in Figure 2.

C. Logistic Regression (LR)

In its basic form logistic regression is a statistical model that uses a logistic function (f(x)=1/(1+e^(-x))) to model the probability that a particular instance falls into a certain class such as pass/fail, healthy/sick, dead/alive and in our case, it is real/fake. In machine learning despite having the word "regression" in its name, logistic regression is used as a parametric classification model.

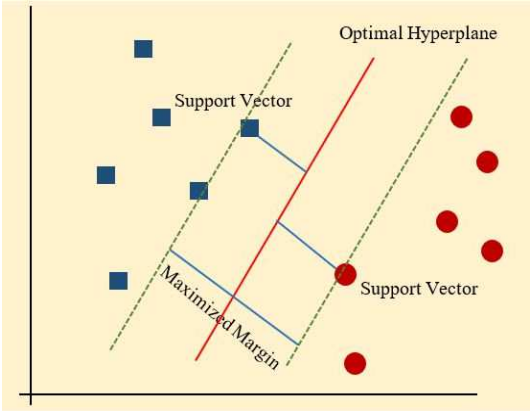

Fig. 2. Support Vector Machine

D. Random Forest (RF)

Random Forest is one of the machine learning algorithms used for classification and regression analysis. The building blocks of a Random Forest are decision trees as represented in Figure 3. A forest is made up of several trees. The central notion behind random forest is a modest but prevailing perception — the wisdom of crowds. Random forest forms decision trees on arbitrarily chosen data instances obtain the predicted output

from each tree and elect the optimal result using a voting mechanism. It is an ensemble method that reduces the over-fitting by averaging the result and is far better than a single decision tree. The parameters used for training Random Forest classifier are max_depth = 45, min_samples_split=7, n_estimators = 100 (denotes no. of decision trees) and random_state = 1.

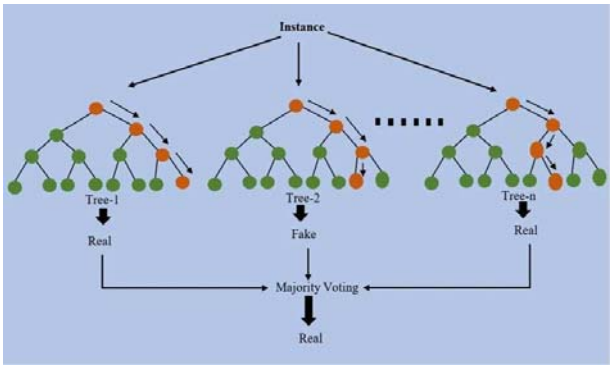

Fig. 3. Random Forest

E. Naïve Bayes (NB)

Naïve Bayes Classifiers is a complete group of algorithms based on "Bayes Theorem" sharing the common principle of mutual independence of every pair of features being classified. Mathematically Bayes theorem is characterized as:

P(A/B) =(P(B/A) P(A))/(P(B)) where
A and B are trials, P(B)≠0
P(A/B) is the likelihood of occurrence of A when B is true.
P(B/A) is the likelihood of occurrence of B when A is true.

P(A) and P(B) are the likelihoods of occurrence of A and B respectively, also known as marginal probabilities. In our fake news detection framework, we specifically use Multinomial Naïve Bayes where the value of parameter alpha=0.1.

F. K-Nearest Neighbor (KNN)

K-Nearest Neighbor also known as KNN algorithm assumes that things that have similar properties exist nearby. KNN uses the idea of similarity which can be sometimes the distance or proximity and can be calculated through various distance calculating algorithms. The majority vote of 'k' nearest data-points are selected amongst all the data-points i.e. classes of k neighbors are considered while classifying the current data point and it is being assigned to the class most common among its k nearest neighbors. We have taken n_neighbors=6, i.e., the value of k=6 so the vote of 6 closest neighbors is considered.

IV. Experiments

To validate the effectiveness of machine learning classifiers we performed experiments on windows 10 64-bit operating system and 8GB RAM with an Intel i5 processor. Programming is done with the Kaggle notebook environment in Python language using online GPU. Performance is evaluated in terms of Accuracy, Recall (Hit rate/Sensitivity), Precision (Positive Predictive Value), F1-score, Misclassification rate (Error rate), Type-I error (False Positive Rate/Fall out) and Type-II Error (False Negative Rate/Miss rate).To get the overall result validation and comparative performance analysis the model was run on two datasets hosted on the Kaggle platform (a) Fake News Detection dataset (FND) and (b) Real or Fake dataset with 80% data used for training and 20% for testing the architecture. The specifics of the datasets are described in Table I.

A. Result Analysis

The performance evaluation of proposed machine learning classifiers on the Fake News Detection (FND) dataset is presented in Table II. The experimental results show that the Support vector machine outperforms all other classifiers by providing 96.61% highest classification accuracy and least error rates. Confusion matrix and ROC curves for SVM are represented in Figure 4 and Figure 5 respectively.

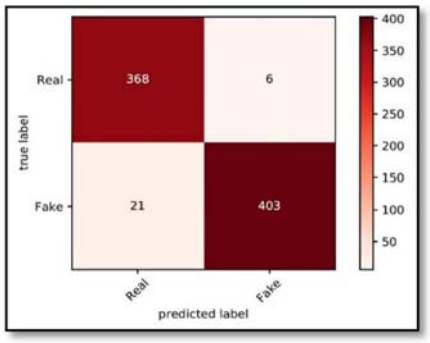

Fig. 4. Confusion Matrix For FND Dataset (SVM Classifier)

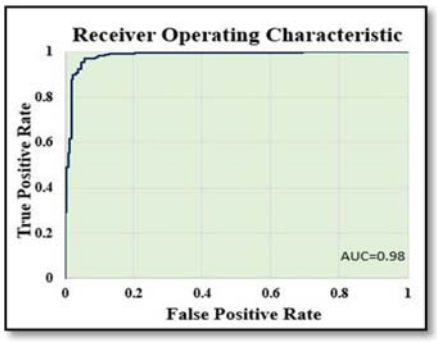

Fig. 5. ROC-AUC Curve For FND dataset (SVM Classifier)

TABLE I. DATASET DETAILS

Sr. No.	Dataset Name	Details	# Initial Instances	# Instances after pre-processing	# Real News	# Fake News
1.	Fake News Detection (Jruvika/FND) [15]	Hosted on Kaggle platform; Contains four fields URLs, Headline, Body, Label	4009	3988	1867	2121
2.	Real or Fake [16]	Hosted on Kaggle platform, contains four fields Id, Headline, Body, Label	6335	6000	3000	3000

TABLE II. RESULT ANALYSIS ON FAKE NEWS DETECTION DATASET

Machine Learned Classifiers	Performance Parameters						
	Accuracy (%)	**Precision (%)**	**Recall/Hit Rate (%)**	**F1-score (%)**	**Misclassification Rate/Error rate (%)**	**Type-I Error/Fall out (%)**	**Type-II Error/Miss Rate (%)**
SVM	**96.61**	**94.60**	**98.39**	**96.45**	**3.39**	**4.95**	**1.61**
LR	91.22	91.30	89.84	94.71	8.78	7.55	10.16
RF	89.47	87.18	90.90	89.00	10.53	11.79	9.10
NB	95.36	92.86	97.59	95.17	4.64	6.60	2.41
KNN	92.10	88.97	94.92	91.85	7.90	10.38	5.08

TABLE III. RESULT ANALYSIS ON REAL OR FAKE DATASET

Machine Learned Classifiers	Performance Parameters						
	Accuracy (%)	Precision (%)	Recall/Hit Rate (%)	F1-score (%)	Misclassification Rate/Error rate (%)	Type-I Error/Fall out (%)	Type-II Error/Miss Rate (%)
SVM	90.42	88.68	92.67	90.63	9.58	11.83	7.33
LR	88.42	87.00	90.33	88.63	11.58	13.5	9.67
RF	79.75	78.84	81.33	80.07	20.25	21.83	18.67
NB	84.83	83.71	86.50	85.08	15.17	16.83	13.50
KNN	80.75	79.33	83.17	81.20	19.25	21.67	16.83

TABLE IV. PERFORMANCE COMPARISON ON FND AND REAL OR FAKE DATASET

Method	Description	Fake News Detection Accuracy (%)	Real or Fake Accuracy (%)
Bali et al. [17]	Random Forest (RF), K-Nearest Neighbour (KNN), Support Vector Classifier (SVC), Multi-Layer Perceptron (MLP), Naïve Bayes (NB), AdaBoost (AB) and Gradient Boosting (XGB) classifiers with GloVe word embedding, cosine similarity, n-gram count and sentiment polarity	86.20	90.02
Agarwalla et al. [18]	Support Vector Machine (SVM), Naïve Bayes (Lidstone smoothing) and Logistic Regression with Punkt statement tokenizer	83.87	88.00
O'Brian et al [19]	CNN learning with word2Vec Embedding and Adam optimizer	94.06	90.00
Reis et al. [20]	KNN, NB, RF, SVM and XGBoost with n-gram, POS tagging, LIWC, Semantic Feature	93.78	89.99
Our Method	SVM with Count Vectorizer and TF-IFD	**96.61**	**90.42**

Table III highlights the performance evaluation of machine learning classifiers on the Real or Fake dataset. The highest classification accuracy with this dataset is 90.42% provided by Support Vector Machine which also demonstrates the least misclassification rate, Type-I error and Type-II error in comparison with other classifiers. Figure 6 and Figure 7 represents the confusion matrix and ROC curves for Real or Fake Dataset using SVM.

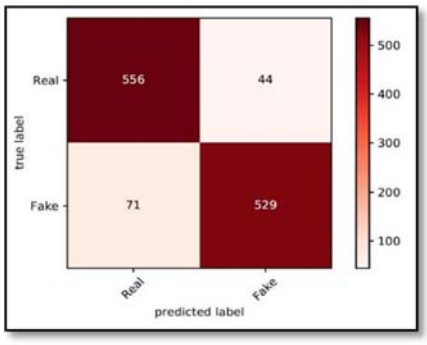

Fig. 6. Confusion Matrix for Real or Fake Dataset(SVM Cassifier)

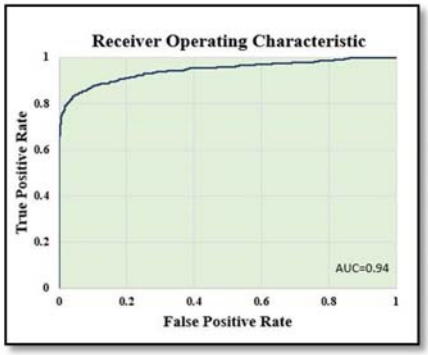

Fig.7. ROC-AUC curve for Reak or Fake Datset (SVM Dataset)

B. State-of-the-art Comparison

The outcomes of the state-of-the-art comparison of the proposed framework with other contemporary methods on identical datasets are being highlighted in Table IV. The reference methods are tested on Fake News Detection and Real or Fake dataset for identical training -testing split (80:20) with the same experimental conditions and the proposed architecture in corresponding research.

V. Conclusion and Future Works

In this work we successfully summarized, compared and analyzed the different aspects of spurious news that go around in social media. The current trend of news consumption, the scenario of fabricated news and state-of-the-art methods for the identification of fake news as well as rumors are being discussed. Five different supervised machine learning models are being trained and tested on two different text news datasets. The highest accuracy 96.61% on the Fake News Detection dataset in classifying fake news using the SVM Classifier looks very promising.

Even though a lot of work has been done by researchers specifically since 2016 to reduce the malice of information pollution from human society still a lot more is to be done. The data could be pre-processed differently; other feature extraction methods such as sentiment analysis, the polarity of the document can also be explored. Proposed work can be further extended to explore the authenticity of the URL which is also present in the dataset. The independent machine learning meta classifiers can be ensembled together through bagging, boosting or voting method. We can try to incorporate the advanced aspect of natural language processing, deep learning and transformers for extracting semantic, linguistic, stylometric hidden patterns in the text of fake writings as well for better detection in the future.

VI. References

[1] K. Sharma, F. Qian, H. Jiang and N. Ruchansky, "Combating fake news: A survey on identification and mitigation techniques," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 10, no. 3, pp. 1-42, 2019.

[2] "United Nations Covid-19 Response," 2020. [Online]. Available: https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19.

[3] Á. Figueira and L. Oliveira, "The current state of fake news: challenges and opportunities," *Procedia Computer Science,* vol. 121, pp. 817-825, 2017.

[4] M. Freeze, M. Baumgartner, ·. P. Bruno, J. R. Gunderson, J. Olin, M. Q. Ross and J. Szafran, "Fake Claims of Fake News: Political Misinformation, Warnings, and the Tainted Truth Effect," *Political Behavior,* pp. 1-33, 2020.

[5] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Systems with Applications,* p. 112986, 2019.

[6] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova and Y. Choi, "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, 2017.

[7] H. Reddy, N. Raj, M. Gala and A. Basava, "Text-mining-based Fake News Detection Using Ensemble Methods," *International Journal of Automation and Computing,* vol. 17, no. 2, pp. 210-221, 2020.

[8] Q. Su, M. Wan, X. Liu and C.-R. Huang, "Motivations, Methods and Metrics of Misinformation Detection: An NLP Perspective," *Natural Language Processing Research,* vol. 1, pp. 1-13, 2020.

[9] Y. Liu and Y.-F. B. Wu, "FNED: A Deep Network for Fake News Early Detection on Social Media," *ACM Transactions on Information Systems,* vol. 38, no. 2, pp. 1-33, 2020.

[10] S. C. R. Gangireddy, D. P, C. Long and T. Chakraborty, "Unsupervised Fake News Detection: A Graph-based Approach," in *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, New Yark, USA, 2020.

[11] H. Q. Abonizio, J. I. d. Morais, G. M. Tavares and S. B. Junior, "Language-Independent Fake News Detection: English, Portuguese, and Spanish Mutual Features," *Future Internet,* vol. 12, no. 5, 2020.

[12] Roohani, T. Rana and P. Meel, "Rumor Propagation: A State-of-the-art Survey of Current Challenges and Opportunities," in *2nd International*

Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur , India, 2019.

[13] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Information Sciences,* vol. 497, pp. 38-55, 2019.

[14] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Computing Surveys ,* vol. 51, no. 2, pp. 1-36, 2018.

[15] [Online]. Available: https://www.kaggle.com/jruvika/fake-news-detection..

[16] [Online]. Available: https://www.kaggle.com/rchitic17/real-or-fake..

[17] A. P. S. Bali, M. Fernandez, S. Choubey, M. Goel and P. K. Roy, "Comparative Performance of Machine Learning Algorithms for Fake News Detection," in *International Conference on Advances in Computing and Data Sciences , Springer* , Singapore, 2019.

[18] K. Agarwalla, S. Nandan, V. A. Nair and D. D. Hema, "Fake News Detection using Machine Learning and Natural Language Processing," *International Journal of Recent Technology and Engineering (IJRTE),* vol. 7, no. 6, pp. 844-847, 2019.

[19] N. O'Brien, S. Latessa, G. Evangelopoulos and X. Boix, "The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors," in *32nd Conference on Neural Information Processing Systems (NIPS 2018)* , Montréal, Canada , 2018.

[20] J. C. S. Reis, A. Correia, F. Murai, A. Veloso and F. Benevenuto, "Supervised learning for fake news detection," *IEEE Intelligent Systems,* vol. 24, no. 2, pp. 76-81, 2019.