IEEE-SEM, Volume 11, Issue 4, April-2023
ISSN 2320-9151

False Message Detection on Social Media: Novel Perspective and Approach

John O. Onigbinde, Alexey N. Nazarov, Ilia M. Voronkov, Alfred O. Blessing and Tuleun T. Daniel.

[1] Department of Cyber Security, Phystech School of Radio Engineering and Computer Technology, Moscow Institute of Physics and Technology (MIPT), Russia.
[2] MIREA – Russian Technological University Moscow, Russia. https://orcid.org/0000-0002-0497-0296
[3] Moscow institute of Physics and Technology (MIPT) National Research University, Russia.
[4] Department of Computer Science, Ahmadu Bello University, Zaria-Nigeria.
[5] Department of Intelligent Information Systems and Technology, Moscow Institute of Physics and Technology (MIPT) National Research University, Russia.
onigbinde.d@phystech.edu , a.nazarov06@bk.ru, alfredblessingogbene@gmail.com

ABSTRACT

Fake messages prove to have a great impact on society as well as the public. It does not only affect people's perception but also fails to preserve the traditional news ecosystem based on the pillars of truth and reality. Considering this situation that affects the public worldwide, here we propose an application that can identify any false information that gets circulated through social media. Our system is proposed with a goal to identify the fake messages by making comparisons with the existing facts and data which Fake messages proves to have a great impact on society as well as the public. The text information given by the user as an input to the system can be easily distinguished either as fake or real with respective tags attached in the output. Our proposed model enables the ability to identify fake and misleading information and thus retain the trust of the public, leading to the protection of society from the negative impacts of fake news available in our datasets. To implement the model, various machine learning ensemble learners were used. The model is trained using an appropriate dataset in Python and performance evaluation was also done using various performance measures. Multi-perceptron neural network binary classifier has the highest accuracy of 96%.

Keywords : Fake messages, Social-Media, Neural Network, Ensemble Learners and Classification

1 INTRODUCTION

IN In the light of the recent attention to the role of social media in the dissemination of fake messages about current political and social affairs, it is important to understand the way the audience interacts with disinformation on Social Network Sites. Fake messages about current social or political issues are circulated on social media with tremendous speed [9]. These fake stories or hoaxes – deliberately or not – misinform or deceive audiences. Usually, these stories are created to either influence people's views, push a political agenda or cause confusion and can often be a profitable business for online publishers. Fake messages can deceive people since their sources are mainly using names and web addresses similar to reputable messages organizations. There are also cases where fake messages are produced by mistake, but they might also confuse and mislead audiences. Many people consume messages and are informed about current political and social affairs from social media platforms and networks and it can often be difficult to tell whether stories are credible or not. Information overload and a general lack of understanding of how the internet works, have also contributed to an increase in fake messages [4]. Both social media and users can play a big part in increasing the spread of these types of stories. However, there are individual users and groups of users who are taking action to counter the spread of fake messages on social media. These groups of people and their actions are the focus of this research so that a more comprehensive framework of how users can identify and fight fake messages can be developed.

Social network organizations like Google and Facebook have announced new measures to tackle fake messages with the introduction of reporting and flagging tools [7]. Media organizations like the BBC and Channel 4 have also established fact-checking sites. While these are positive developments, digital media literacy and developing skills to critically evaluate information are essential skills for anyone navigating the internet. The vast amount of information available online and the rise of fake political messages highlights the need for critical thinking. Therefore, it is crucial to examine the users' acts of verification on spotting and curbing fake messages on social media. The tools and methods they use to identify a fake story, as well as the way they interact with it, can be used to obtain useful information about how users could potentially behave online to counter fake messages on social media.

In addition, fake messages alter the way that individuals connect with genuine messages. A few fake messages are made, intentioned to delude and confuse social media clients, particularly youthful students and old people who are purged of self-protection consciousness. Social media platforms that create fake messages tend to be short-lived. For illustration, numerous dynamic fake messages platforms amid the 2016 U.S. elections did not exist after the campaign [8]. As more consideration is paid to fake messages in recent years, more fake messages generators are nothing but a temporal streak in order to maintain a strategic distance from detection by the detection frameworks. Besides, most

of the fake messages on social media are centering on current events and hot issues to bring more attention to online clients. The real-time nature of fake messages on social media makes distinguishing online fake messages indeed more difficult. It is complicated to assess how numerous online clients are included with a certain piece of instant message, and it is difficult to tell when and how the far-reaching results of fake messages halt. In this research, we propose a model for fake messages detection using machine learning and natural language processing techniques. In particular, we studied and developed methods and tools for detecting fake messages, also proposing a methodology for that purpose and implementing an algorithm that classifies whether the message is fake or real.

2 Related Works

Due to the large number of messages that are transmitted through social media, manual verification is impossible, prompting the development and implementation of autonomous systems for detecting fake messages. There are different approaches proposed by experts to identify the authenticity of online messages.

Shu et al. [11] surveyed the features and models used by detection techniques designed to address fake messages in traditional and social media, considering both message content features and models as well as social context features and models, which can be based on posts, individual users, or user networks, and described the psychological and social foundations of fake messages in traditional and social media.

Alonso et al. [2] distinguished false messages from other kinds of conveying misinformation, such as hoaxes, propaganda, satire/parody, rumors, clickbait, and junk messages. They added misinformation to the list of traditional disinformation and misinformation categories. The spreading of true facts with the aim of damage was defined as misinformation. However, fake and junk messages, which cannot be deemed to contain legitimate information, was cited as a probable source of disinformation, which appears to be contradictory.

Sharma et al. [10] looked at fake message detection and mitigation techniques that rely on computational methods collected a list of available data sets and offered a list of obstacles and outstanding topics. They discovered that sentiment analysis was a useful clue for detecting fake messages, as positive sentiment words in favorable fake reviews tended to be exaggerated compared to their actual counterparts, whereas responses to fake messages on social media tended to be negative.

Zhou et al. [14] looked at fake messages' detection from the perspectives of knowledge-based methods, which check if the knowledge within the text of the message is consistent with facts, style-based methods, which look at how fake messages are written, propagation-based methods, which look at how fake messages spread online, and source-based methods, which look at the credibility of messages sources. They viewed sentiment as a critical semantic component of messages content. They also noted that developing effective and understandable false messages detection algorithms will necessitate collaboration between professionals in computer and information sciences, social sciences, political science, and journalism.

Meel and Vishwakarma [9] investigated the false information ecosystem, from the classification of false information to the incentives for spreading it, as well as the social impact and user perception. They also talked about the state of fact-checking right now, including source detection, propagation dynamics, detection approaches, and containment and intervention mechanisms. They viewed sentiment analysis as one of the most important sources of data for detecting misleading information.

Ahmad et al. [1] extracted linguistic features like n-grams from textual articles and trained multiple machines learning models, including nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), linear support vector machine (LSVM), decision tree (DT), and stochastic gradient descent (SGD). According to the findings, the overall accuracy of a particular article declined as the number of n-grams calculated grew. Learning models that are used for classification have been shown to exhibit this behavior.

The existing literatures do not adequately analyze trained combinations of different machine learning algorithms utilizing various ensemble approaches based on those properties. Ensemble learners have shown to be beneficial in a wide range of applications since the learning models use techniques like bagging and boosting to reduce error rates. These methods make it easier to train various machine learning algorithms in an effective and efficient manner.

3 Methodology

The architectural design in Figure 1 covers the whole procedure from the handling of the dataset to the prediction results. The datasets will be imported using Pandas in Python, Numpy, Sklearn library, genism for the data processing and classification. Matplotlib will be used for visualization of the statistics.

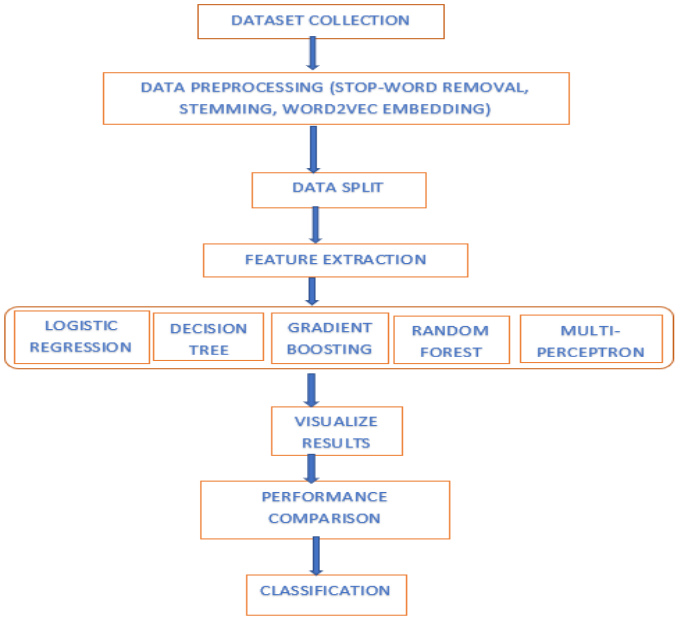

Fig. 1. Model Architecture of Proposed System

3.1 Data Collection

Categorizing a news statement as "fake news" could be a very challenging and time-consuming task. For this reason, the use of an existing dataset becomes imperative. The dataset for this study is from an open source and freely available online. The dataset was collected from the Kaggle. The data includes both fake and real news articles from multiple domains. The real news articles published contain true description of real-world events, while the fake news websites contain claims that are not aligned with facts. The dataset in all contains 44,898 articles among which includes both fake news as well as real news. The fake news data and real news data is separated into two different datasets, with 23,481 fake news articles and 21,417 real news articles. The dataset contains four columns, i.e., title, subject, the main text and the date.

The exploratory data analysis describes the insight of the dataset used in developing the fake news detection model. The dataset contains 23,481 fake news and 21,417 real news. Figure 4.1 shows the distribution of fake news to real news where class "0" represents fake news and class "1" represents real news.

The dataset contains both fake and real news on different subjects. There are 1,570 government news, 778 Middle-east news, 9,050 news, 783 US news, 4,459 left-news, 6,841 politics news, 11,272 politics news and 10,145 world news. The highest number of news is from the politics news while Middle-east news is the lowest. Figure 2 shows the bar chart of fake new and real news. Also. Figure 3 shows the distribution of news per subject.

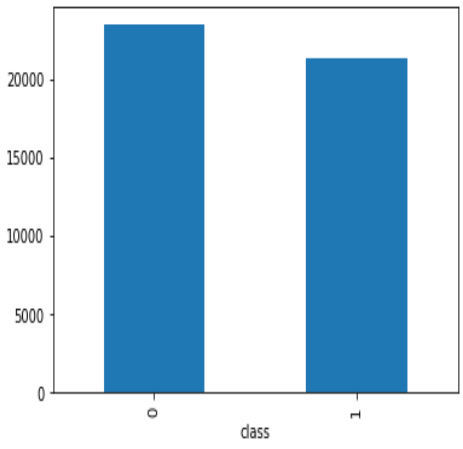

Fig. 2. Bar Chart of Fake News and Real News

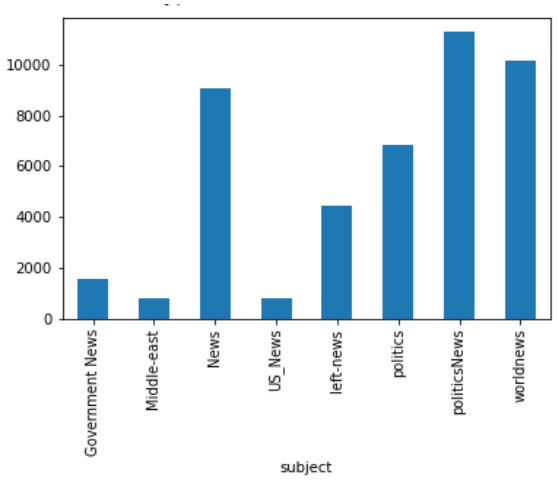

Fig. 3. Distribution of news per subject

3.2 Data Preparation

In creating a vector model, the data must be liable to some refinements such as elimination of stop-words, tokenization, lower case segmentation and removal of punctuations. This helps one to decrease the size of the real data by eliminating the unnecessary data. A standardized feature to eliminate punctuation and non-letter characteristics will be built. High dimension learning is one of the categorization challenges. A large expression number, phrases and words are contained in the data which result in the learning process becoming highly computational. In addition, the classifier precision and output may be influenced by irrelevant and redundant functions. It is better to reduce the feature size and prevent wide area measurements of the feature. The integrations for the classification models will be generated with Word2Vec.

Word2Vec was used for the representation of continuous vector computing of words from larger datasets. It produces faster learning models and is clearly easier. Two predictive models – the CBOW (Continuous Bag of Words) and the Skip-gram were adopted. The predictive models train their vectors to boost their predictive capacities, so that better outcomes are learned.

3.3 Ensemble Learners

Existing ensemble techniques along with textual characteristics as feature input to improve the overall accuracy for the purpose of classification between a true and a fake news were used. Ensemble learners tend to have higher accuracies, as more than one model is trained using a particular technique to reduce the overall error rate and improve the performance of the model. The intuition behind the ensemble modeling is synonymous to requesting opinions of multiple experts before taking a particular decision in order to minimize the chance of a bad decision or an undesirable outcome. A classification algorithm can be trained on a particular dataset with a unique set of parameters that can produce a decision boundary which fits the data to some extent. The outcome of that particular algorithm depends not only on the parameters that were provided to train the model, but also on the type of training data. If the training data contains less variance or uniform data, then the model might overfit and produce biased results over unseen data. Therefore, cross validation is used to minimize the risk of overfitting. A number of models can be trained on different set of parameters to create multiple decision boundaries on randomly chosen data points as training data. Hence, using ensemble learning techniques, these problems can be addressed and mitigated by training multiple algorithms, and their results can be combined for near optimum outcome.

3.4 Model Performance Evaluation

The performance of a classifier may vary based on the size and quality of the text data (or corpus) and also the features of the text vectors. Common noisy words called ‘stop words’ are less important words when it comes to text feature extraction, they don’t contribute towards the actual meaning of a sentence and they only contribute towards feature dimensionality and were discarded for better performance. This helps in reducing the size/dimensionality of the text corpus and add text context for feature extraction. Also, lemmatization is used to convert words to their core meaning and this results in multiple word conversion into a single discrete representation.

The dataset was split into test and train. Python’s “sci-kit learn” library facilitated feature extraction and assortment of selection methods. For feature selection, Common Bag of Words (CBOW) was used. The extracted features are then fed into different classifiers, i.e., logistic regression, decision tree, gradient boosting, random forest and multi-perceptron neural network binary Algorithms from the sci-kit learn. After fitting and training the model, comparison of the ‘f1’ scores is done and confusion matrix is referred, in order to make an educated decision. After fitting all the classifiers, the best performing models was selected as candidate model for fake news classification. Parameter tuning done by implementing Grid Search Algorithm methods on these candidate models is an efficient and reliable approach, and it helps one chose best performing parameters for these classifiers. Finally, the best performing model will be used for prediction task. True positives are the correct predictions of the classifier and false positives are the incorrect predictions. Using these numbers makes the task of calculating precision, recall, F1 scores and accuracy effortless.

Fake news problem is a classification problem that predicts whether a news article is fake or not. The confusion matrix shows the ways in which the classification model is confused when it makes predictions. It gives us insight not only into the errors being made by a classifier but more importantly the types of errors that are being made. These metrics are commonly used in machine learning and enable us to evaluate the performance of a classifier from different perspectives. Specifically, accuracy measures the similarity between predicted fake news and real fake news. Table 1 shows the confusion matrix table.

TABLE 1
CONFUSION MATRIX

Total	Class 1 (Predicted)	Class 2 (Predicted)
Class 1 (Actual)	TP	FN
Class 2 (Actual)	FP	TN

True Positive (TP): when predicted fake news pieces are actually classified as fake news. True Negative (TN): when predicted true news pieces are actually classified as true news. False Negative (FN): when predicted true news pieces are actually classified as fake news. False Positive (FP): when predicted fake news pieces are actually classified as true news.

4 Results and Discussion

This section presents the results of the training models, consisting of the accuracy and the classification metrics of precision, recall, and F1 score. The training results are demonstrated and compared based on each model. The implementation was done using logistic regression classifier, decision tree classifier, gradient boosting classifier, random forest classier and multi-perceptron neural network binary classifier. Classification accuracy was noted for all the models used.

Cross-validation technique was used for splitting the dataset randomly into k-folds. (k-1) folds were used for building the model while kth fold was used to check the effectiveness of the model. This was repeated until each of the k-folds served as the test set. 3- fold cross validation was used for this experiment where 70% of the data is used for training the model and remaining 30% for testing.

The classification accuracy of the logistic regression model is 95%. The confusion matrix is shown in Figure 4. The classification accuracy of the decision tree model is 92%. The confusion matrix is shown below in Figure 5. The classification accuracy of the gradient boosting model is 89%. The confusion matrix is shown in Figure 6. The classification accuracy of the random forest model is 95%. The confusion matrix is shown in Figure 7. The classification accuracy of the multi-perceptron neural network binary model is 96%. The confusion matrix is shown in Figure 8.

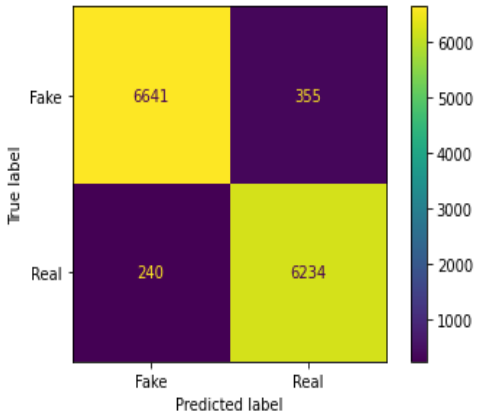

Fig. 4. Logistic Regression Confusion Matrix

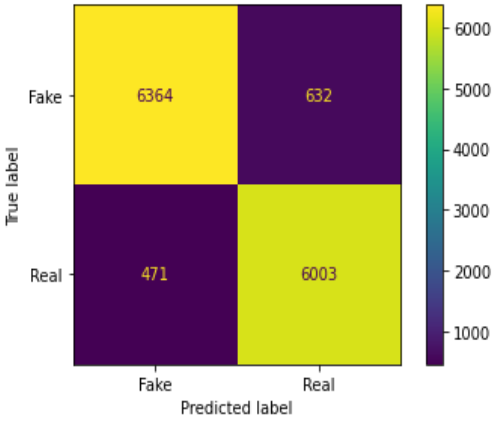

Fig. 5. Decision Tree Confusion Matrix

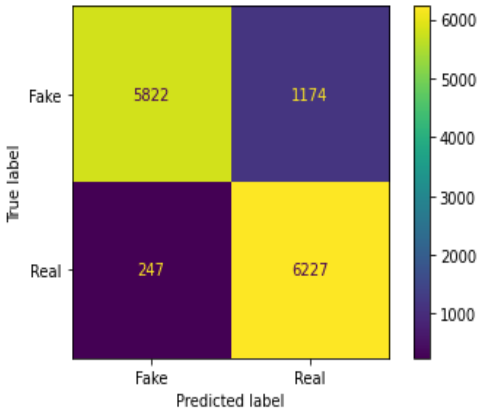

Fig. 6. Gradient Boosting Confusion Matrix

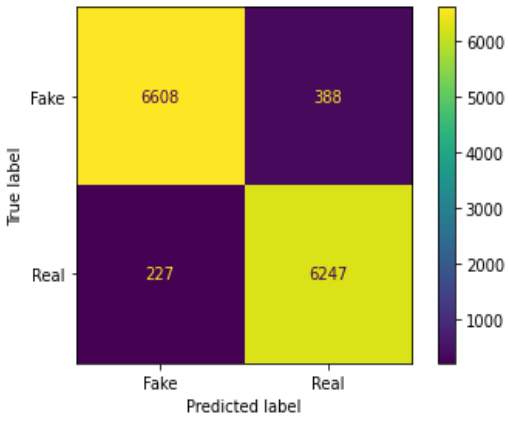

Fig. 7. Random Forest Confusion Matrix

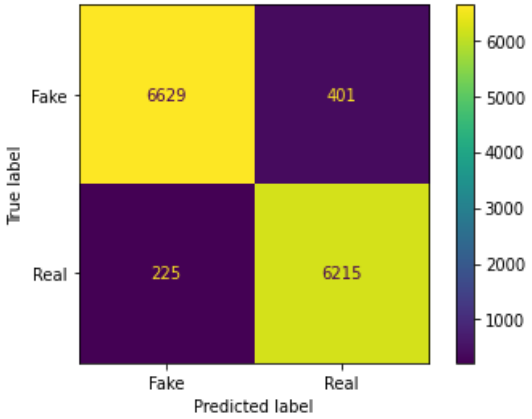

Fig. 8. Multi-Perceptron Neural Network Binary Confusion Matrix

Figure 9 shows the model comparison. Gradient boosting has the minimum accuracy of 89%. Logistic regression and random forest have the same accuracy of 95%, followed by decision tree with 92% accuracy. Also, from the graph, multi-perceptron neural network can be seen to outperform the remaining models with an accuracy of 96%.

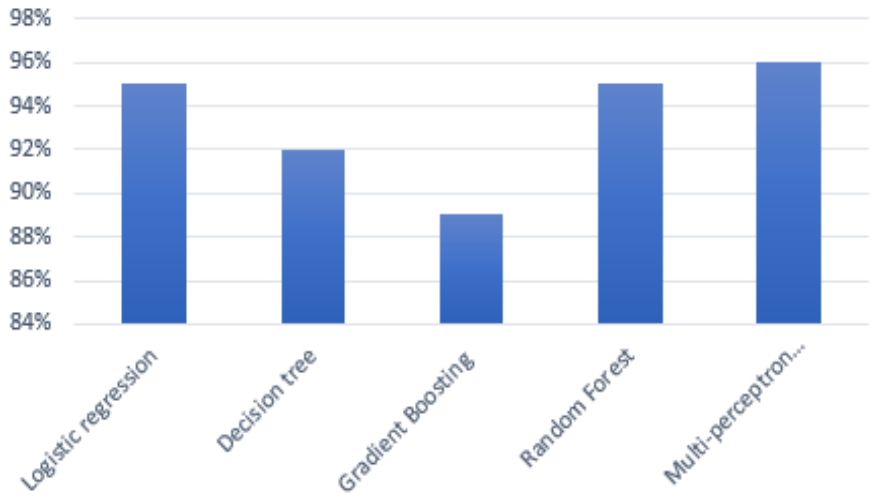

Fig. 9. Comparison of Model Accuracy

Fake news classification using machine learning, deep learning and natural language processing (NLP) was implemented. NLP methods, such as Tokenize, TF-IDF and Word2vec, were applied to increase the accuracy. The five machine learning classifiers used to implement the models are — logistic regression classifier, decision tree classifier, gradient boosting classifier, random forest classier and multi-perceptron neural network binary classifier. The novel empirical studies from previous works were to classify tagged false news, and NLP is integrated with classical machine learning to create an AI model. From the experiments, we can see that the multi-perceptron neural network binary classifier achieves excellent accuracy compared to the other models. As evident from the results, our best model came out to be multi-perceptron neural network binary classifier with an accuracy of 96%.

5 CONCLUSION

The task of classifying news manually requires in-depth knowledge of the domain and expertise to identify anomalies in the text. In this research, we discussed the problem of classifying fake news articles using machine learning models and ensemble techniques. The data we used contains news articles from various domains to cover most of the news rather than specifically classifying political news. The primary aim of the research is to identify patterns in text that differentiate fake articles from true news. The learning models were trained and parameter-tuned to obtain optimal accuracy. Some models have achieved comparatively higher accuracy than others. We used

multiple performance metrics to compare the results for each algorithm. Multi-perceptron neural network can be seen to outperform the remaining models with an accuracy of 96%.

Fake news detection has many open issues that require attention of researchers. For instance, in order to reduce the spread of fake news, identifying key elements involved in the spread of news is an important step. Graph theory and machine learning techniques can be employed to identify the key sources involved in spread of fake news. Likewise, real time fake news identification in videos can be another possible future direction. Dataset is to be extended and more knowledge from an emotional standpoint is to be explored. Lexicon-based documents with language-impartial procedures in a multilingual manner can be explored.

REFERENCES

[1] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M.O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. Complexity, 2020, 1-11.

[2] Alonso, M.A., Vilares, D., Gómez-Rodríguez, C., & Vilares, J. (2021). Sentiment Analysis for Fake News Detection. Electronics, 10(11), 1348.

[3] Baines, D., & Elliott, R.J. (2020). Defining Misinformation, Disinformation and Malinformation: An Urgent Need for Clarity During the COVID-19 Infodemic. Discussion papers, 20(06), 20-06.

[4] Bermes, A. (2021). Information Overload and Fake News Sharing: A Transactional Stress Perspective Exploring the Mitigating Role of Consumers' Resilience During COVID-19. Journal of Retailing and Consumer Services, 61, 102555.

[5] Conroy, N.K., Rubin, V. L., & Chen, Y. (2015). Automatic Deception Detection: Methods for Finding Fake News. Proceedings of the association for information science and technology, 52(1), 1-4.

[6] Golbeck, J., Robles, C., & Turner, K. (2011). Predicting Personality with Social Media. In CHI'11 extended abstracts on human factors in computing systems (pp. 253-262).

[7] Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance. Big Data & Society, 7(1), 2053951719897945.

[8] Majó-Vázquez, S., Congosto, M., Nicholls, T., & Nielsen, R.K. (2021). The Role of Suspended Accounts in Political Discussion on Social Media: Analysis of the 2017 French, UK and German Elections. Social Media+ Society, 7(3), 20563051211027202.

[9] Meel, P., & Vishwakarma, D.K. (2020). 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, 153, 112986.

[10] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating Fake News: A Survey on Identification and Mitigation Techniques. ACM Transactions on Intelligent Systems and Technology (TIST), 10(3), 1-42.

[11] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

[12] Sivakumar, S., Videla, L.S., Kumar, T.R., Nagaraj, J., Itnal, S., & Haritha, D. (2020, September). Review on Word2vec Word Embedding Neural Net. In 2020 international conference on smart electronics and communication (ICOSEC) (pp. 282-290). IEEE.

[13] Zhang, X., & Ghorbani, A.A. (2020). An Overview of Online Fake News: Characterization, Detection, and Discussion. Information Processing & Management, 57(2), 102025.

[14] Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.