International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Approach towards Fake News Detection using Machine Learning Techniques

Mr. Vyankatesh Rampurkar[1], Dr. Thirupurasundari D.R.*[2]

Submitted: 26/01/2024 **Revised**: 04/03/2024 **Accepted**: 12/03/2024

Abstract: In the digital age, the spread of false information has become a widespread and difficult problem. The Naive Bayes & logistic regression algorithms are used in this paper to provide a novel methodology for the detection of bogus news stories. The aim of this study is to improve the efficacy of the identification of fake news in digital material, consequently fostering information credibility and integrity within the digital ecosystem. We start this investigation by gathering a wide dataset of news articles from both reputable and phoney sources. We preprocess the textual input using techniques like tokenization, stop-word removal, and stemming to aid in feature extraction. During the feature selection phase, the term frequency-inverse document frequency (TF-IDF) is used to estimate the word importance of each article. Next, the Naive Bayes algorithm is used to divide news stories into two groups: phoney and real. In order to determine the probability that an article will fall into a particular category, Naive Bayes uses a probabilistic technique under the assumption that the characteristics (words) are conditionally independent. Logistic Regression models the probability of a news article being fake or genuine based on a set of relevant textual features. The primary goal of logistic regression is to achieve high accuracy in classifying news articles as fake or genuine, with an emphasis on feature engineering and model evaluation. The efficacy of the corresponding methods is determined by utilizing the confusion matrix to evaluate the correctness of the model. The findings suggest that Logistic Regression is effective in detecting fake news and contributes to the trustworthiness of information sources in the digital age.

Keywords: *Confusion Matrix, ISOT Dataset, Machine Learning, Navie Bayes, Logistic Regression*

1. Introduction

Information that is inaccurate or misleading but is presented as fact is called "fake news." It can appear in many different forms, such as biased reporting that misrepresents the facts, rumours, manufactured stories, altered photos or videos, and hoaxes. Fake news is frequently produced with the goal of misleading or deceiving the public and is typically distributed through a variety of media platforms, such as social media, websites, and traditional news publications. False or misleading information can be disseminated by fake news, which can cause people to believe and do untrue things. When individuals start to doubt the veracity of the information they come across, the spread of fake news has the potential to undermine confidence in institutions, media, and even in each other. Fake news has the potential to exacerbate polarization and conflict by taking advantage of already-existing political and social gaps. Misinformation regarding catastrophes, health crises, or other significant occurrences can cause harm or bad decision-making among the public. Fake news has the potential to cause economic instability, harm company reputations, and have an impact on financial markets. Election manipulation, public opinion manipulation, and democratic process undermining are all possible with the deployment of fake news.

Detecting and addressing fake news is therefore essential for maintaining the integrity of information in the digital age and for safeguarding the public against misinformation and its negative consequences. Detecting fake news helps prevent people from being misled and making decisions based on false information that could be harmful. Identifying and debunking fake news helps to preserve trust in credible news sources and institutions. Detecting fake news can help mitigate the polarization and social division that can result from its spread.

Therefore objectives to detect fake news are as follows:

Early detection: Identifying fake news as quickly as possible to limit its spread and mitigate its impact.

Fact-checking: Verifying the accuracy of information by comparing it to reliable sources and evidence.

Identifying sources: Determining the credibility of the sources that share the information.

Analysing content: Assessing the content of the news for signs of manipulation, bias, or inconsistency.

Automation: Developing automated tools and algorithms to assist in the rapid detection of fake news at scale, especially on social media platforms.

In the digital age, the spread of false information has become a widespread and difficult problem. The Naive Bayes algorithm, a popular machine learning method for

[1] Bharath Institute of Higher Education and Research, Chennai
ORCID ID : 0000-0002-8515-3713
[2] Bharath Institute of Higher Education and Research, Chennai
** Corresponding Author Email: harshaldada@gmail.com*

text classification tasks, is used in this paper to provide a novel methodology for the detection of bogus news stories. The aim of this study is to improve the precision and efficacy of the identification of fake news in digital material, consequently fostering information credibility and integrity within the digital ecosystem. We start this investigation by gathering a wide dataset of news articles from both reputable and phoney sources. To help with feature extraction, we preprocess the textual input using methods including tokenization, stop-word removal, and stemming. Term frequency-inverse document frequency (TF-IDF) is used in the feature selection process to measure the significance of words in each article. Next, news stories are divided into two categories using the Naive Bayes algorithm: phoney and authentic. Naive Bayes uses a probabilistic method to calculate the probability that an article falls into a particular category, based on the assumption that the characteristics (words) are conditionally independent. Confusion matrix is used to measure the accuracy of the model and determine the algorithm's efficacy.

This paper also presents a targeted approach for fake news detection through the utilization of Logistic Regression, a powerful machine learning algorithm. Logistic Regression models the probability of a news article being fake or genuine based on a set of relevant textual features. A comprehensive dataset, comprising textual features extracted from news articles, forms the foundation of this study. The primary objective of this research is to achieve high accuracy in classifying news articles into the categories of fake and genuine. The findings of this study underscore the effectiveness of Logistic Regression as a valuable tool for fake news detection.

2. Literature Review

To detect fake news, numerous scholars have proposed applying a range of machine learning and deep learning methods. We have provided several basic methods for detecting bogus news in this research study. Finding the flaws in the baseline methodology and offering a In a research publication [1], the effectiveness of several Machine Learning and Deep Learning models is evaluated using two different datasets. Text representation is derived by the application of embedding techniques such as term frequency and term frequency-inverse document frequency. Term frequency is utilized for text representation in machine learning methods, and term frequency-inverse document frequency is used for deep learning models. Subsequently, performance metrics such as accuracy, precision, and F1-score are employed to compare different models. Next, the unique stacking process is proposed as an improved strategy to detect fake news. By stacking the several models, the proposed approach enhances the performance of each model. This work compares the performance variations among individual models and enhances diversity by training three deep learning models as Convolutional neural network, Long Short-Term Memory, Gated recurrent unit and five machine learning algorithms as Decision tree, Random forest ,k-nearest neighbours, Logistic regression and Support vector machine. In research paper [2], Rumor Detection on Social Networks using a Sociological Approach is discussed. The quick spread of rumors through networks makes distinguishing between rumors and non-rumors extremely challenging. It is impossible to keep up with the entire worldwide stream of tweets in real-time due to the size of the real-time twitter feed and the rapid arrival rate. Alternately, the proposed model is referred to as a tweet rumor scoring system based on the four ratings produced from the tweet content. It defines four different scores namely, i) Famousness score, ii) Rareness score, iii) Reliability score and iv) Consistency score. To examine the total model classification accuracy, four common machine learning techniques are proposed: K-nearest neighbor (kNN), Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes. The Random Forest (RF) classifier was found to be the most accurate for rumour categorization goals. Ozbay et al. [31] used 23 supervised AI algorithms on three datasets, including BayesNet, JRip, OneR, Decision Stump, ZeroR, Stochastic Gradient Descent (SGD), Logistic Model Tree (LMT), etc. Their testing results demonstrated that the decision tree approach outperformed all other intelligent classification algorithms, with the exception of recollection. Random Forest, Multinomial Naïve Bayes, Gradient Boosting, Decision Tree, Logistic Regression, and Linear-SVM were employed by Kaliyar et al. [21] in their detection of bogus news. They discovered that gradient boosting yields cutting-edge outcomes and on the Fake News Challenge dataset, they were 86% accurate. Umer et al. [40] suggested a model to detect whether headlines of news items coincide with the body of the article by combining dimensionality reduction techniques, PCA and Chi-Square, with neural network architecture, including CNN and LSTM. They then saw that the suggested model produced the best accuracy, 97.8%, in a significantly shorter amount of time. Using the Kaggle fake news dataset, Kaliyar et al. [22] developed a CNN-based deep neural network known as FNDNet, which produced state-of-the-art results with an accuracy of 98.36%. Using the CNN BiLSTM ensemble model with attention mechanism, Kumar et al. [24] obtained the greatest accuracy of 88.78% using both their own datasets and the FakeNewsNet dataset. Ajao et al. [4] classified fake news messages from Twitter tweets using a CNN and RNN hybrid. They evaluated the effectiveness of three different models: the LSTM-CNN hybrid, the LSTM with dropout regularisation model, and the plain LSTM model. The dataset they used comprised about 5,800 tweets that were focused on five rumours. They then came to the conclusion that the LSTM-CNN hybrid model and the LSTM approach with dropout regularisation both suffer from limited data, while the plain LSTM model performs best. In order to classify, Roy et al. [36] used CNN to find hidden features, RNN to record temporal sequence, and MLP to receive the obtained representation. They obtained feature embeddings using pre-trained 300-dimensional Google News Vectors and input them into different convolutional layers and different Bi-LSTM layers. Their models outperform the state-of-the-art model by 3%, with an accuracy of 44.87% on the Liar Dataset.

3. Materials and Methods

3.1. System Architecture

System Architecture to detect fake news using machine learning approach is as shown in figure. The fundamental concepts of proposed models have been discussed in below sections.

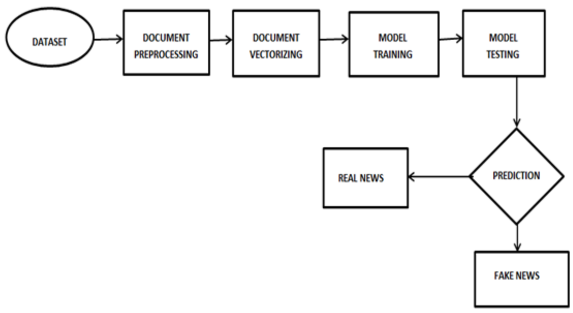

Fig 1. System Architecture to detect fake news using machine learning approach

3.2. Dataset Collection

We have used an ISOT dataset. 23481 pieces of data in this dataset are tagged as fake news (labelled as 0), whereas 21417 pieces of data are classified as legitimate news (labelled as 1). Among the features are the title, subject, date, label and substance (news body). We used the title and text from the ISOT dataset, among other variables, to train our models.

3.3. Document Preprocessing

The text data has to be preprocessed using techniques including stop word removal, tokenization, sentence segmentation, and punctuation removal before being input into machine learning models. These actions can greatly aid in improving model performance and helping us choose the most pertinent terms. Since the dataset is derived from actual news items, many of the links in it are useless and provide no information. Thus, we eliminated these urls from our data in order to clean it. The next stage in our preprocessing is to remove stop words. In case they made too much noise, we took them out. Following the cleaning process, we used TF-IDF and embedding algorithms to tokenize the text data. TF and TF-IDF are two important concepts in the field of information retrieval and natural language processing used for text analysis and document retrieval.

3.4. Document Vectorizing

Documents will get vectorized using TF-IDF and proper score will be assigned to each term in the document.

3.4.1 TF (Term Frequency):

Term frequency, or TF for short, is a measure of how frequently a term (or word) appears in a text. It is calculated by counting the occurrences of a term in a document and is typically normalized

to prevent bias towards larger documents. One common normalization technique is to divide the term frequency by the total number of terms in the document. The frequency factor (TF) of a term in a text is frequently given as a number; larger values indicate that a term appears more frequently in that document.

TF (t, d) = count of t in d / number of words in d

3.4.2 IDF (Inverse Document Frequency):

IDF stands for "Inverse Document Frequency." It is a statistical measure used in natural language processing and information retrieval to evaluate a term's importance or uniqueness within a collection of documents. IDF is a crucial component of the TF-IDF.

$$IDF(t) = \log(N / (n\_t + 1))$$

Where, N is the total number of documents in the corpus.

n_t is the number of documents containing the term "t."

3.4.3 TF-IDF:

A statistical metric called TF-IDF combines TF and IDF to assess a term's significance in a particular document in relation to a broader corpus of texts. It is computed by multiplying the IDF of a term throughout the whole document collection by the TF of the term inside a document. The value of a phrase within a text is expressed by the TF-IDF score in relation to its relevance throughout the full corpus of documents. Excessive TF-IDF values signify that a phrase is unique to a certain document, suggesting that it is common in one text but rare in another.

TF-IDF (t, d) = TF (t, d) * IDF (t)

3.5. Model Training & Validation

In this study, Naive Bayes and logistic regression these machine learning models have been utilized to identify fake news from authentic news. There has been extensive discussion about these models.

3.5.1. Navie Bayes

In machine learning, the straightforward and popular Naive Bayes classification technique is utilized. It is especially well-suited for text classification and spams filtering and is based on the Bayes theorem. The premise that features are conditionally independent that is, that the existence of one feature does not influence the existence of another gives rise to the word "naive" in its name. The foundation of Naive Bayes is the Bayes theorem, which determines the likelihood of a class or hypothesis depending on the likelihood of the supporting data (or features). This is one way to express Bayes' theorem:

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)}$$

Where, P (C|X) is the posterior probability of class C given evidence X.

P (X|C) is the likelihood of evidence X given class C.

P(C) is the prior probability of class C.

P(X) is the probability of evidence X.

Naive Assumption: Naive Bayes relies on the "naive" assumption that characteristics, such as words in a text document, are conditionally independent based on the class. Because each feature's likelihood may be treated separately, this makes the computation simpler.

Training: To train a Naive Bayes classifier, you estimate the prior probabilities (P(C)) and the likelihood (P (X|C)) for each class (C) using training data. In text classification, this typically involves counting the frequency of each word or feature in documents for each class.

Prediction: To make a forecast for a new set of data, you take the class with the highest likelihood out of each class's posterior probability. When classifying news, for instance, in news prediction, you compute the posterior probabilities of the two classes (fake and not fake) and assign the news to the class with a greater likelihood.

3.5.2. Logistic Regression

A statistical and machine learning model called logistic regression is utilized for multiclass classification with certain adjustments as well as binary classification. Regression modelling is used to forecast the likelihood that an input will belong to a specific class. The salient features of logistic regression are as follows:

Binary Classification: In binary classification, where the output variable might have one of two possible values (e.g., 0 or 1, yes or no), logistic regression is mainly utilized. It simulates the relationship between the likelihood that an input will belong to the positive class and a set of independent factors, or features.

Sigmoid Function: The logistic function, commonly known as the sigmoid function, is used in logistic regression to model the probability.

$$S(z) = \frac{1}{1+e^{-z}}$$

Where, a linear combination of the characteristics and model parameters is represented by 'z' in this instance.

Linear Combination: The linear combination of the model parameters and features in logistic regression is computed as follows:

$z=\beta_0+\beta_1 x_1+ \beta_2 x_2+\ldots..+ \beta_n x_n$

Where, β_0, β_1, $\beta_2 \ldots \beta n$ are the coefficients or model parameters & $x_1, x_2, \ldots x_n$ are the features

Log-Odds: It is commonly known as the log-odds or logit, which is the result of the linear combination 'z'. It is the natural logarithm of the probability that the event will occur.

Probability Estimation: The log-odds 'z' is mapped to a probability value in the interval [0, 1] via the sigmoid function. The following formula determines the likelihood that an occurrence is in the positive class:

$P(Y=1|X) = S(z)$

Model Training: Using methods like maximum likelihood estimation, the model parameters (β_0, β_1, β_2,...) are calculated from the training data. In light of the model, these parameters are changed to maximize the likelihood of the observed data.

Decision Boundary: A decision boundary is used by logistic regression to categorize fresh data points. Since the threshold is set by default at 0.5, an instance is classified as belonging to the positive class if the anticipated probability is greater than or equal to 0.5, and as belonging to the negative class otherwise.

When forecasting one of two choices, logistic regression is used similarly to a tool for decision-making. Let's say you want to find out if a news article is real or not. Logistic regression analyses word and phrase features to calculate the chance (probability) of the news being real. If the likelihood is more than fifty percent, it declares something to be true. If it is less than 50%, it says, "No, it's not real." As a result, it facilitates binary decision-making using the data at hand.

4. Result Analysis

Result analysis gives a brief summary of both Naive Bayes and Logistic Regression's capacities for identifying false news. The study's findings indicate that while Logistic Regression outperforms Naive Bayes in terms of accuracy and precision, both models are useful in identifying false news. Both models showed strong F1 scores and accuracy, indicating that they can distinguish between real and fraudulent news. Both models had a comparatively low amount of false positives, according to the confusion matrices, which is encouraging for preventing the incorrect classification of authentic news as fraudulent. The performance of the models is assessed using accuracy, precision, recall, F1 Score & Confusion Matrix. The following outcomes were attained when the models were assessed using the ISOT dataset.

Classification report for Navie Bayes is as shown in table 1. Confusion matrix produced by Navie Bayes algorithm is as shown figure 2. Classification report for Logistic Regression is as shown in table 2. Confusion matrix produced by Logistic Regression algorithm is as shown figure 3. The graphical representation of accuracy produced by navie bayes & logistic regression is as shown below. Logistic Regression outperformed Naive Bayes in every evaluation criterion, proving its supremacy in identifying fake news. The high recall of both models indicates their effectiveness in identifying false positives, which is crucial in lowering false negatives. The results of the investigation show that both models are helpful in spotting false news, even though Logistic Regression performs better than Naive Bayes in terms of accuracy and precision. The model selection is influenced by specific needs, computational resources, and the significance of different performance metrics. Future studies may look into deeper learning models, ensemble approaches, or more advanced machine learning techniques to further improve the capacity to identify bogus news.

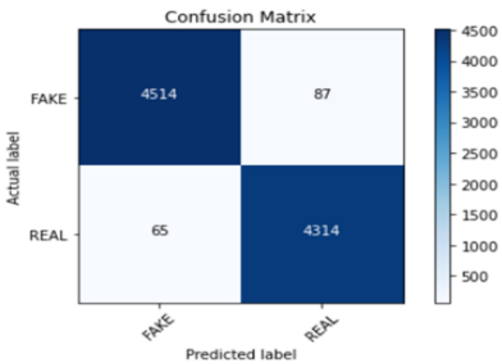

Fig 2. Confusion matrix produced by Navie Bayes

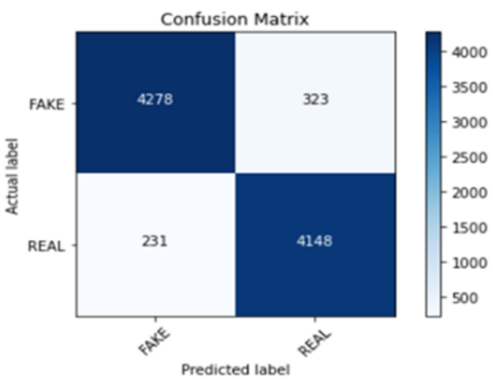

Fig 3. Confusion matrix produced by Logistic Regression

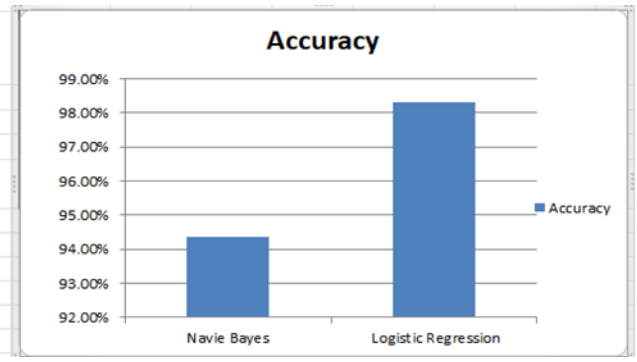

Fig 4. Accuracy comparison between Navie Bayes & Logistic Regression

Table 1. Classification report for Navie Bayes

Navie Bayes	**Precision**	***Recall***	***F1 Score***
Fake	0.96	0.94	0.95
Real	0.93	0.95	0.94
Accuracy	94.37%		

Table 2. Classification report for Logistic Regression

Logistic Regression	**Precision**	***Recall***	***F1 Score***
Fake	0.99	0.98	0.98
Real	0.98	0.99	0.98
Accuracy	98.31%		

5. Conclusion

In this study, we evaluated the performance of two machine learning models, Naive Bayes and logistic regression, using the ISOT dataset. Our analysis concludes that logistic regression is the preferred choice for fake news detection due to its impressive accuracy of 98.31%. Logistic regression outperforms Naive Bayes in several aspects, especially in the context of modern information dissemination. It provides greater flexibility, enabling the capture of complex relationships and dependencies within the data, which is crucial for distinguishing between fake and genuine news, given their diverse linguistic and contextual cues. Additionally, logistic regression is more robust against the curse of dimensionality and excels in handling high-dimensional data compared to Naive Bayes. Its ability to manage complex correlations and utilize regularization techniques makes it a more effective tool for combating fake news in our information-driven society.

Acknowledgements

Authors like to thank to Prof. Thirupurasundari D.R. who is working as an Associate Professor at Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, India (Chennai) for her fruitful suggestions during the preparation of the manuscript.

Author contributions

Conceptualization: Mr. V.V. Rampurkar

Methodology: Mr. V.V. Rampurkar

Investigation: Mr. V.V. Rampurkar.

Discussion of results: Mr. V.V. Rampurkar

Writing – Original Draft: Mr. V.V. Rampurkar

Writing – Review and Editing: Mr. V.V. Rampurkar.

Resources: Mr. V.V. Rampurkar

Supervision: Dr. D.R. Thirupurasundari

References

[1] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapè, ''MIMETIC: Mobile encrypted traffic classification using multimodal deep learning,'' Comput. Netw., vol. 165, Dec. 2019, Art. no. 106944.

[2] G. E. R. Agudelo, O. J. S. Parra, and J. B. Velandia, ''raising a model for fake news detection using machine learning in Python,'' in Proc. Conf. e- Bus., e-Services e-Soc. Cham, Switzerland: Springer, 2018, pp. 596–604..

[3] H. Ahmed, I. Traore, and S. Saad, ''Detecting opinion spams and fake news using text classification, '' Secur. Privacy, vol. 1, no. 1, p. e9, Jan. 2018.

[4] O. Ajao, D. Bhowmik, and S. Zargari, ''Fake news identification on Twitter with hybrid CNN and RNN models, '' in Proc. 9th Int. Conf. Social Media Soc., Jul. 2018, pp. 226–230.

[5] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, and A. Gelbukh, '''Bend the truth': Benchmark dataset for fake news detection in urdu language and its evaluation, '' J. Intell. Fuzzy Syst., vol. 39, no. 2, pp. 2457–2469, 2020.

[6] S. D. Bhattacharjee, A. Talukder, and B. V. Balantrapu, ''Active learning based news veracity detection with feature weighting and deep-shallow fusion, '' in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2017, pp. 556–565.

[7] C.-C. Chang and C.-J. Lin, ''LIBSVM: A library for support vector machines, '' ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, p. 27, 2011.

[8] H.-L. Chen, B. Yang, J. Liu, and D.-Y. Liu, ''A support vector machine classifier with rough set-based feature selection for breast cancer diagno- sis, '' Expert Syst. Appl., vol. 38, no. 7, pp. 9014–9022, Jul. 2011.

[9] Kakarwal, Sangeeta, and Pradip Paithane. "Automatic pancreas segmentation using ResNet-18 deep learning approach." System research and information technologies 2 (2022): 104-116.

[10] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernelbased Learning Methods. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[11] M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro, and L. de Alfaro, ''Automatic online fake news detection combining content and social signals, '' in Proc. 22nd Conf. Open Innov. Assoc. (FRUCT), May 2018, pp. 272–279.

[12] A. L. Edwards, ''Note on the 'correction for continuity' in testing the significance of the difference between correlated proportions, '' Psychome- trika, vol. 13, no. 3, pp. 185–187, 1948.

[13] P. H. A. Faustini and T. F. Covões, ''Fake news detection in multi- ple platforms and languages, '' Expert Syst. Appl., vol. 158, Nov. 2020, Art. no. 113503.

[14] S. Gilda, ''Evaluating machine learning algorithms for fake news detection, '' in Proc. IEEE 15th Student Conf. Res. Develop. (SCOReD), Dec. 2017, pp. 110–115.

[15] M. H. Goldani, S. Momtazi, and R. Safabakhsh, ''Detecting fake news with capsule neural networks,'' 2020, arXiv:2002.01030.

[16] Paithane, Pradip Mukundrao. "Yoga Posture Detection Using Machine Learning." Artificial Intelligence in Information and Communication Technologies, Healthcare and Education: A Roadmap Ahead 27 (2022).

[17] A. U. Haq, J. Li, M. H. Memon, J. Khan, S. U. Din, I. Ahad, R. Sun, and Z. Lai, ''Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for predic- tion of Parkinson disease,'' in Proc. 15th Int. Comput. Conf. Wavelet Act. Media Technol. Inf.

Process. (ICCWAMTIP), Dec. 2018, pp. 101–106

[18] Paithane, Pradip, Sarita Jibhau Wagh, and Sangeeta Kakarwal. "Optimization of route distance using k-NN algorithm for on-demand food delivery." System research and information technologies 1 (2023): 85-101.

[19] A. Jain, A. Shakya, H. Khatter, and A. K. Gupta, ‘‘A smart system for fake news detection using machine learning,’’ in Proc. Int. Conf. Issues Challenges Intell. Comput. Techn. (ICICT), vol. 1, Sep. 2019, pp. 1–4.

[20] R. K. Kaliyar, A. Goswami, and P. Narang, ‘‘Multiclass fake news detection using ensemble machine learning,’’ in Proc. IEEE 9th Int. Conf. Adv. Comput. (IACC), Dec. 2019, pp. 103–107.

[21] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, ‘‘FNDNet—A deep convolutional neural network for fake news detection, ’’ Cognit. Syst. Res., vol. 61, pp. 32–44, Jun. 2020

[22] S. Kula, M. Choraś, R. Kozik, P. Ksieniewicz, and M. Woźniak, ‘‘Sentiment analysis for fake news detection by means of neural networks, ’’ in Proc. Int. Conf. Comput. Sci. Cham, Switzerland: Springer, 2020, pp. 653–666.

[23] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, ‘‘Fake news detection using deep learning models: A novel approach, ’’ Trans. Emerg. Telecommun. Technol., vol. 31, no. 2, p. e3767, Feb. 2020

[24] Y. Long, ‘‘Fake news detection through multi-perspective speaker profiles, version I17-2043,’’ Assoc. Comput. Linguistics, Strouds- burg, PA, USA, Tech. Rep., Nov. 2017, vol. 2. [Online].

[25] W. Y. Wang, ‘‘‘Liar, liar pants on fire’: A new benchmark dataset for fake news detection,’’ 2017, arXiv: 1705.00648.