



Fake News Detection Using Artificial Neural Network Algorithm

U. J. Isha Priyavamtha, G. Vishnu Vardhan Reddy^(✉), P. Devisri,
and Asha S. Manek^(✉)

Department of CSE, RV Institute of Technology and Management, Bangalore, India
gadusunuryvishnur_cs19.rvitm@rvei.edu.in

Abstract. As technology is increasing rapidly, nowadays most people prefer social media over newspapers for the news. There are chances of spreading fake news widely on social media because of circulation of news can be done by anyone. The fake news can show adverse effects on economic as well as political issues. It can spread hatred among the community which can cause serious problems to peace in society. Hence, it is important to detect fake news. The developed model using Artificial Neural Network is a solution for the problem of fake news detection and filtered to avoid wide spread. The proposed model detects fake news that includes three main phases 'preprocessing, feature extraction and classification'. Initially input is preprocessed to extract features using clustering algorithms. Subsequently, the model is developed to detect fake news. The Neural Networks and Linear Support Vector Clustering algorithms resulted in 99.90% and 97.5% accuracy respectively. Finally, analysis is carried out for validating the betterment of the presented model in terms of different measures.

Keywords: Artificial Neural Network · Fake News Detection · Preprocessing · Support Vector Clustering

1 Introduction

When newspapers, television and radios were the only source for the news, only authorized people were involved in preparing news [1]. Now-a-days, most of the news is being escalated on internet social community rather than by the news channels or newspapers as it is easy to access and cost effective. This made the spread of fake news much easier. The fake news is spreading at a very large rate, for example, during "U.S presidential election", 2016, more than 25% of Americans has visited a fake news website in the span of six weeks. Fake news is usually created and spread for economic and political benefits [2]. It will show adverse effects when communal hatred is spread by means of fake news. One of the best examples for the destruction caused by the fake news is the "Delhi riots" [1]. Fake news is misleading information that circulates across the internet as news. The rapid escalation of fake news in internet social community is resulting in loss of public trust in social media. Hence, there is an immense need for news detection

and classification of the news as fake news and genuine news in today's scenario to prevent the spread of fake news. It is also necessary to understand that what news contents come under fake news and what does not come under fake news for detection.

There are three types of fake news:

- 1) **CLICKBAIT:** News which displays a sensational headline which is created to obtain more "clicks" which in turn adds to ad revenue.
- 2) **SPONSORED CONTENT:** News which appears as genuine news but in fact it is news for advertising.
- 3) **FABRICATED JOURNALISM:** News which is completely created for economic or political benefits.

The news which won't fall under fake news category is: satire, reporting mistakes and the news which people won't like (claim actual truth also as "fake" just because people don't agree with it). Over half of the population claims that they regularly come across fake news on social media like Facebook, Twitter. In 2018, at least 20 people were killed as a result of fake news which was spread through social media platforms.

In this work, Machine Learning algorithms i.e. Neural Networks algorithm is used to detect fake news. This algorithm outperformed with 99.90% accuracy whereas Support Vector Clustering algorithm results 97.5% accuracy and Passive Aggressive Classifier gives 96.67% accuracy.

1.1 Research Goal

The objective of the work is to examine the fake news and filter out from the social media containing false and misleading information. Primarily, training data from Kaggle website news.csv [3] is preprocessed and used for training the model. The data from news.csv is then categorized into three characteristics (i) authenticity (ii) intention and (iii) benign news. Further news has been categorized into clickbait, sponsored content and fabricated journalism. Analysis is made based on these characteristics of the news to classify into fake and benign news for purpose of helping users to avoid being lured by clickbait and fabrication.

1.2 Our Contribution

The contribution of the proposed solution is as follows:

- The proposed research work discusses the potential directions to improve fake news detection and reduction capabilities.
- Owing to this, the ANN model is developed that can identify and remove fake news from the results provided to a user by internet social community or a search engine.
- The proposed model provides future directions to detect fake news in social internet media.
- The proposed model performs a binary classification on a news.csv dataset and classifies it into "real" or "fake".

The structure of the research paper is as follows: Sect. 2 shows a literature review of the paperwork. Section 3 discusses the datasets and explains the implementation of the methods for fake news detection. Section 4 depicts the obtained results and compares the results with other existing methods. Section 5 concludes the paper.

2 Literature Survey

2.1 Related Work

Aman Srivatsava [1] has built a website for fake news detection composite of Machine Learning and NLP methods. A combination of five Machine Learning techniques used to build the website are ‘Support Vector Machine (SVM)’, ‘Logistic Regression’, Naive Bayes, Random Forest Classifier, Stochastic Gradient Descent (SGD) in combination with Bag of Words, TF-IDF, POSTAG and N-GRAM feature selection. After comparing the accuracy and F1-Score of the model with different combinations of techniques, author concluded the paper with 95% accuracy and 0.95 F1-Score using SVM classifier in combination with TF-IDF and POSTAG.

Waikhom et al., used the publicly available LIAR dataset for fake news detection collected from POLITIFACT.COM. The LIAR dataset provides links to source documents for each case. The work provides ensemble techniques for fake news detection [2].

Xinyi Zhou et al., has used four perspectives for fake news detection surveys based on knowledge, style, propagation and source methods. The survey also highlights potential research tasks based on the news reviews and identify as fundamental theory across various disciplines to encourage interdisciplinary research on fake news [4].

Nicole O’Brien has used Machine Learning technology for the detection of fake news. The accuracy of 95.8% obtained by the model for the dataset of 5,504 fake and real news. The classification of news was based on language patterns using Neural Network classifier [5].

Sharma, K et al., described current problems facing due to fake news in survey paper [6] by focusing the associated technical challenges. The authors focused on the advancements of each method along with advantages and limitations.

2.2 Review

The pros and cons of the existing models regarding the fake news detection system is illustrated in Table 1. Still, at present, the Machine Learning classifier is used in most applications, where it can get impacted by various outer sources. The Artificial Neural Network (ANN) is used in a few and is in the infant stage. Hence more research is deployed in the proposed ANN model. Even the used classifier algorithms have some disadvantages that need a further extension. Some of the pros and cons are explained as follows:

Coh-Matrix-feature extraction tools [7] outperforms with 92% classification accuracy by satisfying readability features considering the Brazilian Portuguese language to detect fake news. But this methodology has to deepen the study on syntactic and semantic features.

Data augmentation technique is used to expand the limited annotated resources for Urdu by applying machine translation from English to Urdu. However, needs improvement in expanding annotated resources for other language pairs with larger parallel resources [8].

WeFEND framework [9] works by integrating three components including the annotator, the reinforced selector and the fake news detector on WeChat dataset consisting

Table 1. Features and challenges of existing models regarding the fake news detection.

Author [citation]	Methodology	Features	Challenges
Roney L. S. Santos et al. [7]	Coh-Matrix-feature extraction tools used for extracting readability metrics	- Analysis of readability features influence on the Brazilian Portuguese language fake news - 92% classification accuracy	Need to deepen study on syntactic and semantic features
Maaz Amjad et al. [8]	Data augmentation technique to expand the limited annotated resources for Urdu	- Investigated whether machine translation from English to Urdu can be applied as a text data augmentation technique to expand the limited annotated resources for Urdu - Findings indicate that character bigrams are the most robust and highest performing feature for the datasets augmented with machine-translated texts	Need to investigate whether text data augmentation through Machine Translated (MT) dataset for other language pairs with larger parallel resources will provide more promising results
YaqingWang et al. [9]	WeFEND framework	- The proposed framework works by integrating three components including the annotator, the reinforced selector and the fake news detector on WeChat dataset consisting of news articles and user feedback with 82.4% accuracy	Need to improve model performance by finding the optimum value for each parameter

of news articles and user feedback with 82.4% accuracy. But still needs betterment in performance accuracy and further methods to improve efficiency.

3 Methodology and Machine Learning Techniques

3.1 Data Collection

The printed and digital media information authenticity is a major concern which badly affects businesses and society. The information reaches and affects with fast pace and gets amplified, distorted, inaccurate on social media. Then this false information acquired enormous potential to cause bad impact on large users in real world. The existing datasets [3] of news articles annotated with veracity is used as a basis for this experimental study. The training dataset `trainnews.csv` contains six columns and 25116 rows. The description of each of the columns is given in Table 2.

3.2 Preprocessing

The preprocessing is used to process the dataset which consist of noise, missing values and redundancy to obtain a clean dataset. The size of the dataset is reduced because of elimination of unnecessary elements such as over representation of words and repeated occurrences in datasets.

Table 2. Description of the column of dataset [3]

Sl. no.	Column name	Description
1	id	Unique Identification Number of each news article
2	headline	The title of the news
3	news	The full text of the news article
4	serial_no	A serial number
5	written_by	The author of the news article
6	label	Labels - news is fake (1) or not fake (0)

Data is preprocessed at different levels as follows:

- The rows with 'NaN' values are removed from the dataset.
- Label Encoding is used to convert labels into integer values.
- The function 'train_test_split' is imported from 'sklearn.model_selection' to split the entire dataset into testing and training dataset.
- The news data is converted into a CSR sparse matrix based on the Term Frequency - Inverse Document Frequency (TFIDF) Vectorizer.
- The data is converted into 'NumPy' float array.
- This array is given to the input layer of the Artificial Neural Network.

The preprocessed dataset obtained by applying separate python code to the dataset [3] and steps performed are as shown in Fig. 1 below.

3.3 Feature Selection

Term Frequency- Inverse Document Frequency (TF-IDF) method is generally used to find frequent occurring words in classification of documents because it is easy, uncomplicated and high level of processing rate of weighing feature method.

TF-IDF: TF-IDF is a calculation of authenticity of a word by differentiating the number of times a word occurs in a document with the number of documents the word occurs.

$$TF - IDF = TF(\text{term, document}) \times IDF (\text{term}) \tag{1}$$

Where,

TF-Term Frequency: Number of times term t appears in a document d.

IDF - Inverse Document Frequency is given by:

$$IDF(\text{term}) = \log \left[\frac{(1 + n)}{(1 + df(d, t))} \right] \tag{2}$$

Where,

N - Number of documents.

df (document, term) - Document frequency of the term t.

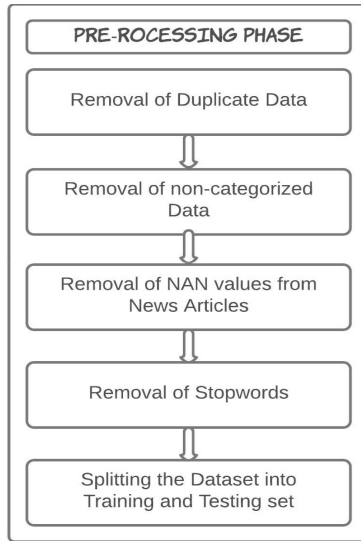


Fig. 1. Steps performed during pre-processing phase.

3.4 Evaluation Metrics

In order to evaluate classification models, accuracy is the one evaluation metric used. The accuracy is defined as the fraction of correct predictions of the proposed model.

$$Accuracy = \frac{\text{Number of corrected predictions}}{\text{Total number of predictions}} \tag{3}$$

3.5 Models

Training, preprocessing and post processing are performed in cloud based environment called as ‘Collaboratory’ i.e. COLAB. The Pandas library, ‘Jupyter Notebook’ and the Python programming languages were used to develop the proposed model. The flow diagram of the proposed automatic fake news detection model is as shown in Fig. 2.

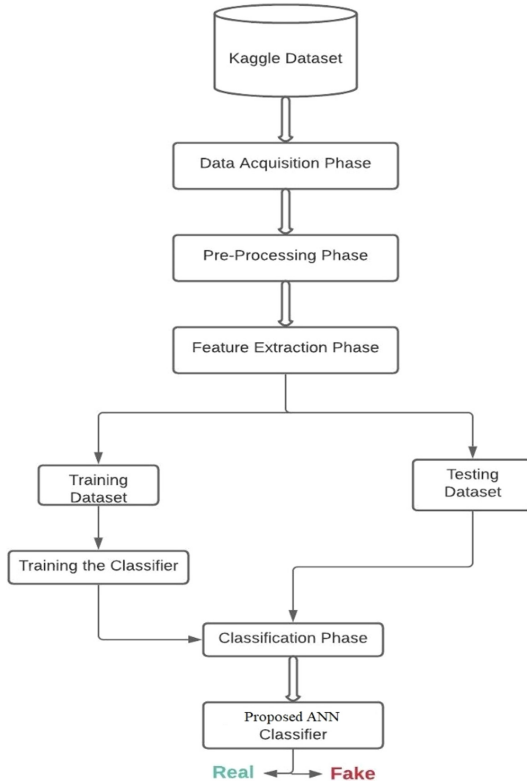


Fig. 2. The flow diagram of the proposed automatic fake news detection system.

Classification problems can be solved by Machine learning in two steps:

1. Learning the model from a training data set.
2. Classifying the hidden data on the basis of the trained model.

The way of organizing Machine learning algorithms is based on the desired result of the algorithm or the type of input available during training of the machine. The following are few of the Machine learning algorithms used for fake news detection and classification of articles for selected dataset [3].

- (a) **Logistic Regression:** Logistic Regression algorithm is used to predict the probability of a categorical dependent variable. In Logistic Regression, the dependent

variable is a binary variable that contains data coded as 1 (yes, real, etc.) or 0 (no, fake, etc.). In other words, the Logistic Regression model predicts $P(Y = 1)$ as a function of X . The obtained accuracy for the model is 94.4% with Liblinear solver approach as shown in Fig. 3.

- (b) **Passive Aggressive Classifier:** Passive Aggressive Classifier (PAC) algorithm learns from the model by staying passive i.e., do nothing if prediction is correct else aggressive i.e. model tries to correct the prediction. The obtained accuracy for the model is 95.15% as shown in Fig. 4.
- (c) **Naïve Bayes:** Naïve Bayes algorithm is a probabilistic classifier based on Bayes Theorem. Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. The obtained accuracy for the model is 79.95% as shown in Fig. 5.

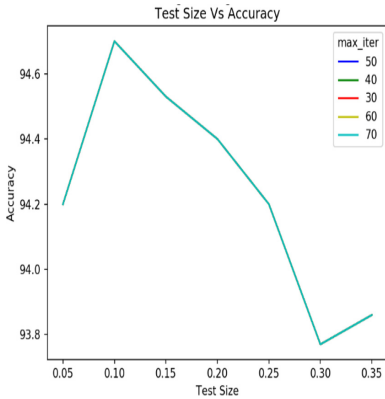


Fig. 3. Accuracy v/s test size in Logistic Regression algorithm.

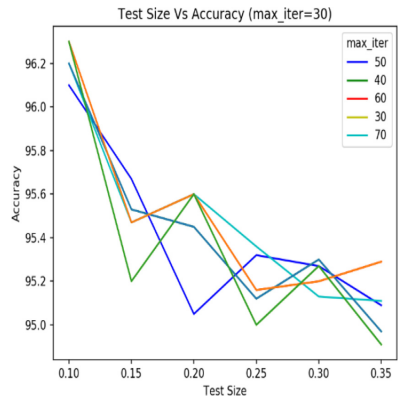


Fig. 4. Accuracy v/s test size in Passive Aggressive Classifier.

- (d) **Stochastic Gradient Descent Classifier:** Stochastic Gradient Descent (SGD) algorithm is a simple yet very efficient approach to fit linear classifiers and regressors under convex loss functions such as (Linear) Support Vector Machines and Logistic Regression. The obtained accuracy for the model is 93.8% as shown in Fig. 6.
- (e) **Linear Support Vector Clustering (Linear SVC):** The purpose of a Linear SVC is to fit to the data provided, returning a “best fit” hyper plane that divides, or categorizes the provided data. After getting the “best fit” hyper plane, some features are fed to the classifier to check for predicted class.

In SVC model, data points are outlined from data space to a high dimensional feature space using a Gaussian kernel. The smallest sphere enclosing the image of

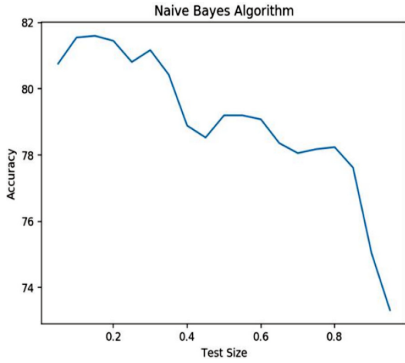


Fig. 5. Accuracy v/s test size in Naive Bayes algorithm.

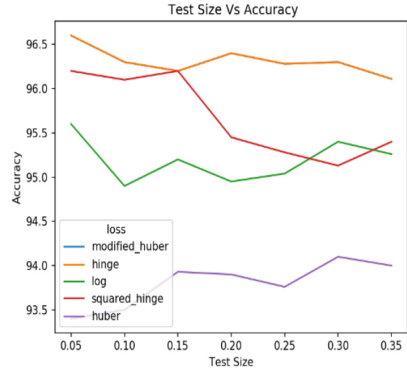


Fig. 6. Accuracy v/s test size in Stochastic Gradient Descent algorithm.

the data is considered in feature space to map back to data space forming a set of contours called as cluster boundaries enclosed by data points. Points enclosed by each separate contour are associated with the same cluster. The width parameter of the Gaussian kernel decreases resulting in the increased number of clusters with increase in number of disconnected contours in data space. The contours can be interpreted as delineating the support of the underlying probability distribution. The model gives the accuracy 97.5% with split size of 90% training and 10% testing dataset. The obtained graph of test size v/s accuracy is as shown in Fig. 7.

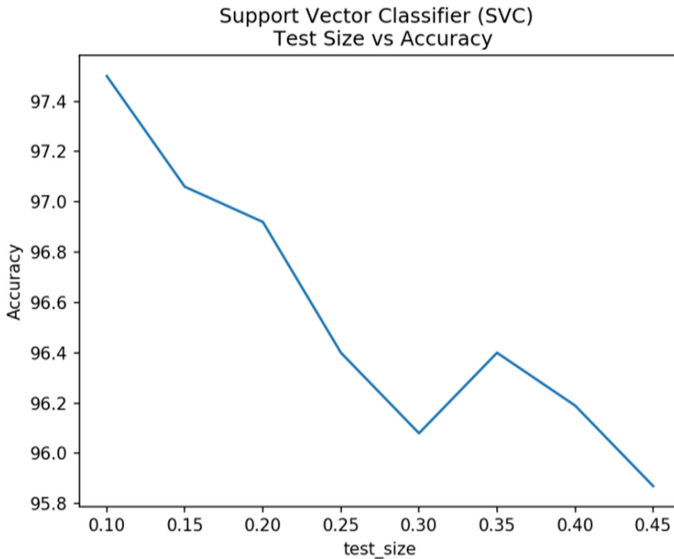


Fig. 7. Accuracy v/s test size in linear SVC algorithm

- (f) **Artificial Neural Network (ANN):** Neural networks are the sequence of algorithms which mimic the functioning of the human brain to make predictions. In Neural Networks, the input layer computes the data and transfers it to the output layer to make predictions. The loss of the prediction is calculated by loss function algorithm ‘binary cross entropy’ which is used for binary classification models. Gradient descent algorithm results in decreasing the loss and directs the loss function towards global minima. The new weights will be updated in the nodes by back propagation as shown in Fig. 8.

The design of Artificial Neural Network architecture and specification of hyper parameters need to be set before training the model. The Table 3 below shows the hyper parameter values which are set before training the model.

In fake news detection, the dataset is label encoded to convert labels into binary data. The whole news text is vectorized using a TFIDF Vectorizer which assigns the values based on statistical weight age of word. The NumPy array input is fed to the Neural Network as input to make predictions.

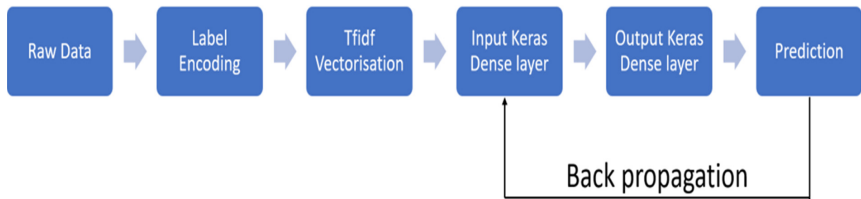


Fig. 8. Back propagation algorithm

The model gives the accuracy 99.90% and loss graphs of Neural Networks in fake news prediction is as shown in Fig. 9.

Table 3. Hyperparameter values of ANN

Name of the hyperparameter	Hyperparameter value
Learning rate	0.001
Batch size	64
Maximum number of epochs	40

While analyzing the relation between accuracy in both training and testing datasets, initially found the model is overfitting over the training dataset. So, added a dropout layer to the Neural Networks to reduce the over fitting.

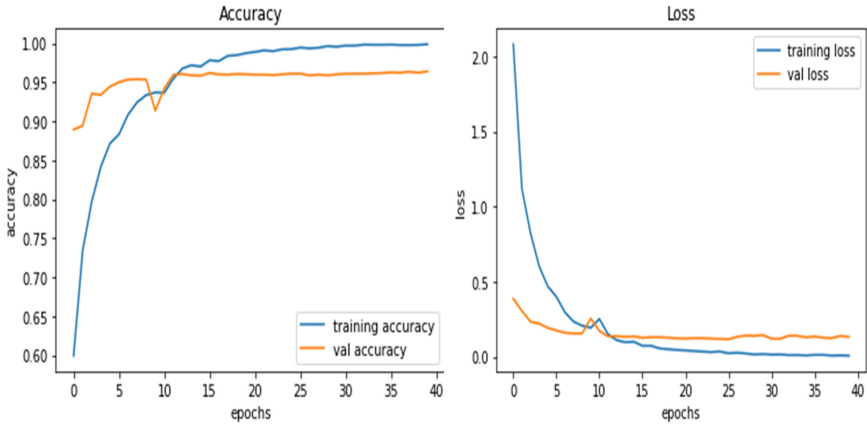


Fig. 9. Accuracy graph and loss graph of Neural Networks

4 Results and Performance Analysis

All trained models for the news dataset in proposed research work outperformed with enough accuracy which can be applied practically for detecting fake news. The performance is depicted in Table 4 are best proved with Linear SVC and Artificial Neural Network model with an accuracy of 97.5% and 99.90% accuracy respectively.

4.1 Performance Evaluation

The dataset is randomly divided into split ratio of 80:20 for training and testing datasets respectively with five-fold cross-validation, among which Linear SVC and ANN performed best compared with the other classifiers (e.g., Logistic Regression; Passive Aggressive Classifier, Stochastic Gradient Descent and Naïve Bayes).

The performance of the classifier is measured in terms of accuracy and result is presented in Table 4.

Table 4. Results of the proposed model for selected Kaggle dataset [3]

Classifier	Accuracy (%)
Logistic Regression (LR)	94.4
Passive Aggressive Classifier (PAC)	96.67
Naïve Bayes (NB)	79.95
Stochastic Gradient Descent (SGD)	96.88
Linear SVC	97.5
Artificial Neural Network (ANN)	99.90

(i) *Comparison of the Proposed Model with Work [10].*

The results show that the proposed model is slightly outperformed in comparison with other similar models for fake news early detection. The proposed Linear SVC and ANN classifiers reports 97.5% and 99.90% accuracy respectively as compared with the XGBoost and Random Forest classifier in work [10] with 89.2% and 84.5% accuracy respectively as depicted in Table 5.

Table 5. Comparison of the proposed model with work [10]

Classifier	Work [10]		Proposed model	
	XGBoost	Random Forest	Linear SVC	ANN
Accuracy	89.2%	84.5%	97.5%	99.90%

(ii) *Comparison of the Proposed Model with Work [11].*

In the proposed research work [11], results obtained by using Neural Network and LSTM classifiers are 93% and 95% accurate with the dataset fetched from Kaggle. As compared to the work proposed in [11], the proposed model outperformed with 97.5% and 99.90% accuracy with Linear SVC and ANN classifier respectively as depicted in Table 6.

Table 6. Comparison of the proposed model with work [11]

Classifier	Work [11]		Proposed model	
	Neural network with Kears	LSTM	Linear SVC	ANN
Accuracy	93%	95%	97.5%	99.90%

5 Conclusion

Fake news data has been categorized into three characteristics as (i) authenticity (ii) intention (iii) benign news. Further it has been categorized into (i) click bait (ii) sponsored content (iii) fabricated journalism. Fake news makes things more complicated and causes a change in people’s mind about digital technology. The proposed model concludes by implementing fake news detection model with python coding. The proposed Artificial Neural Network model fits to the data provided. Passive Aggressive Classifier learns staying passive and Naïve Bayes Classifier uses probabilistic classification Bayes theorem. Stochastic Gradient Descent is a very efficient approach in fitting linear classifiers and regressors under Support Vector Machine and Logistic Regression.

Comparison of all the models is tabulated showing the accuracy levels. The proposed model's accuracy score is 99.90% and obtained by applying various NLP and Machine Learning techniques.

The research can be further carried out for various news datasets by using unsupervised machine learning classifiers. The proposed model can be improved by increasing the dataset size.

References

1. Srivastava, A.: Real time fake news detection using machine learning and NLP. *Int. Res. J. Eng. Technol. (IRJET)* **07**(06) (2020)
2. Waikhom, L., Goswami, R.S.: Fake news detection using machine learning. In: *Proceedings of International Conference on Advancements in Computing & Management (ICACM)* (2019)
3. <https://www.kaggle.com/surekharamireddy/fake-news-detection>
4. Zhou, X., Zafarani, R.: A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Comput. Surv. (CSUR)* **53**(5), 1–40 (2020)
5. O'Brien, N.: Machine learning for detection of fake news. Ph.D. dissertation, Massachusetts Institute of Technology (2018)
6. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y.: Combating fake news: a survey on identification and mitigation techniques. *ACM Trans. Intell. Syst. Technol. (TIST)* **10**(3), 1–42 (2019)
7. Santos, R., et al.: Measuring the impact of readability features in fake news detection. In: *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 1404–1413 (2020)
8. Amjad, M., Sidorov, G., Zhila, A.: Data augmentation using machine translation for fake news detection in the Urdu language. In: *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 2537–2542 (2020)
9. Wang, Y., et al.: Weak supervision for fake news detection via reinforcement learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 516–523 (2020)
10. Zhou, X., Jain, A., Phoha, V.V., Zafarani, R.: Fake news early detection: a theory-driven model. *Digit. Threats Res. Pract.* **1**(2), 1–25 (2020)
11. Pavan, M.N., Prasad, P.R., Gowda, T., Vibhakar, T.S., Shidnal, S.: Fake news detection using machine learning. *Int. Res. J. Eng. Technol. (IRJET)* **07**(12), 2151–2154 (2020)