

Fake News Detection Using Supervised Machine Learning Classification Algorithms



Akanksha Singh and Sanjay Patidar

Abstract The online world has become an essential part of everybody's life in today's society. Almost everyone is using social media to gain information about all around the world. The news or information shared on the Internet could be beneficial as well as harmful. On one hand, it is the most inexpensive, easy, and convenient way of getting information in no time; on the other hand, it also prevails fake news. False news/information has a tremendous impact on our social lives, in fact, in all fields, particularly politics and education, and organization. The propagation of false information has the potential to create significant social and emotional harm, as well as have potentially dangerous consequences. Spreading incorrect data via online media to stand out enough to be noticed or monetary and political increase is common on social media these days. The focus of the research is to develop a detection that incorporates multiple machine learning classification methods in order to present an analytical method for detecting false news. To identify fake news, we used five machine learning classification approaches. The five supervised ML classifiers are logistic regression, decision tree, Naïve Bayes, Random Forest, and support vector machine. We have calculated the accuracy of different classifiers and gave a comparative analysis of accuracy along with all performance measures. The model's output has a 99.59% accuracy when employing feature extraction approaches like term frequency and inverted document frequency, and a decision tree as a classifier.

Keywords Fake news · Logistic regression · Decision tree · Naïve Bayes · Random forest · Support vector machine

1 Introduction

For a long time, social media has taken over a meaningful place in people's life. Fake news primarily prevails via social media and articles available online. Fake news indulges politics, democracy, education as well as finance and business at risk.

A. Singh · S. Patidar (✉)

Department of Software Engineering, Delhi Technological University, New Delhi, India
e-mail: sanjaypatidar@gmail.com

Even while false news is not a new issue, people these days place a larger focus on social media, which leads to the acceptance of deceitful remarks and the subsequent propagation of the same wrong information. It's getting harder to tell the difference between accurate and misleading news these days, which leads to confusion and complications. Manually recognizing fake news is tough; it is only achievable when the individual identifying the news has extensive expertise in the subject. Fake news can destroy someone's career and if it is political and harm the nation and citizens of that country as well as it can also affect businesses, products, and reputations. It is now easier to manufacture and circulate fake news because of the recent advances in computer science, but it is much more difficult to determine if the information is accurate or not.

As a result, we carried out and compared five various methods in this study. Here, we are using supervised ML classification techniques, i.e., logistic regression, decision tree, Naïve Bayes, Random Forest, and support vector machine to determine if the news being transmitted is authentic or not. We have used a dataset available on the Kaggle website which contains two datasets for there are both authentic and fraudulent news items.

The paper structure could be defined here given there are Sect. 5 in this paper. In Sect. 1, we have provided the formal introduction of the research being carried out and its motive. In Sect. 2, we have discussed the relevant research that has been done in this field and the algorithm that we used in this project. Then in Sect. 3, we've gone through the methodology, which includes information on the flowchart, dataset, and machine learning techniques that we employed in this research. In Sect. 4, we have discussed the implementation and the results obtained in this study. After that in Sect. 5, we have concluded the study so far and the results with accuracy. Also, we proposed the future work for the study.

2 Literature Survey

The primary goal of this study is to discover the best effective classification system for detecting and quantifying false news. To find out, we looked at several classification techniques and used them in our model. We have applied five classification techniques here. Further, we are providing a brief review of the papers we have studied.

In 2017, Granik and Mesyura presented a methodology for detecting bogus news utilizing Naive Bayes on news posts on Facebook and got a 74% accuracy rate. And concluded that AI techniques could be used successfully to handle these kinds of problems [1].

In 2017, Ahmed, Traore, and Saad have suggested a methodology for identifying false news, researchers have developed an AI model using n-gram analysis. The classifier used here was the support vector machine, which has a precision of 92% [2].

In 2017, Campan et al in their study proposed a model how fake news spread on social media and how the Internet affects the diffusion of false information in creating and spreading. They also discussed the solutions to reduce the dissemination of false information and provided the future research aspects in this area [3].

In 2017, Perez-Rosas et al, Klienberg, and colleagues suggested a model that automatically detects fake news for online resources. They developed a computer algorithm and tools to detect bogus news. They work with two different datasets. The first came via the Internet, and the second resulted from a mix of human data collecting and Internet assistance [4].

In 2018, Aphiwongsophon and Chongstitvatana have a study on using Naïve Bayes, SVM, and neural networks to detect the fake news and calculated the performance measures they have found that Naïve Bayes has 96.08% and neural network and SVM 99.90% accuracy. Through this experiment, they found out that neural networks and support vector machines are having significant accuracy and high confidence [5].

In 2018, Gahirwal and colleagues suggested a support vector machine news detection based model for false or real news that has an accuracy of 87%. She had recognized comedy, negative words, ridiculousness, syntax, and punctuation using five predictive features. Its goal was to ensure that the substance of a news piece was accurate [6].

In 2019, Ozbay and Alatas have used AI techniques for detecting fake news. In the first phase, they preprocessed the dataset to transform unstructured data into structured data, and then they used text mining to construct about twenty-three supervised AI algorithms. They applied these algorithms to about three real-world datasets and found the accuracy and performance measures accordingly. The best average value they got was by using a decision tree, ZeroR, CVPS, and WIHW algorithms [7].

In 2019, Agarwal et al and colleagues taken a dataset namely the Liar dataset, and given a comprehensive study of various approaches. They offer a stacking model in this research that fine-tunes the informational knowledge obtained from user input at each level before attempting to predict something [8].

In 2019, Riece et al are working on looking for a range of elements in news articles, postings, and stories that might assist to identify false news with increased precision. He demonstrated the significance of these new qualities in evaluating bogus news. Discrimination, integrity, involvement, domain location, and temporal patterns are some of these characteristics. They used 2282 BuzzFeed items in their analysis (news articles). Using KNN, Naive Bayes, Random. Forest, XGBoost, support vector machine (SVM), and they analyzed and described the strengths and limits of this technology and discovered that XGBoost performed better when compared with other with an accuracy of 0.86 [9].

In 2020, Zhang and Ghorbani's study elaborates that false information has been a serious concern for the industry as well as academia as it is widely utilized to confuse and persuade online users with skewed facts. Furthermore, the Internet generates and disseminates a vast volume of fantastic and incorrect information. It has emerged as a potential danger to social networking groups and has had a major adverse influence on online activities such as online commerce and networking sites [10].

In 2020, Shaikh and Patil have used three AI techniques for giving the detection model for the detection of false news. Three algorithms that they used were SVM, Naïve Bayes, and passive aggressive classifier, respectively, with SVM giving the highest accuracy of 95.05% [11].

In 2020, Smitha and Bharath have illustrated the model and different methodologies to identify and quantify fake news with the help of ML techniques and NLP techniques. Seven different classification algorithms are proposed here, and accuracy, F1score, recall, and precision are compared [12].

In 2020, Kesarwani et al demonstrated a basic strategy for detecting false news on social media using a K -nearest neighbor classifier, which obtained an accuracy of roughly 79% when evaluated against a sample of Facebook news articles [13].

In 2021, Khanam et al and colleagues carried out the research by reviewing it in two phases: firstly, they used multiple supervised learning algorithms to define the essential principles and criteria of false news found in web-based media. They proposed using scikit-learn library for processing text data. They performed techniques for feature selection to select the best fit [14].

In 2021, Nagaraja and colleagues showed in their study that false information mostly circulates through social media and is propagated further without investigating the true data. They applied various NLP techniques and two ML algorithms, i.e., Naïve Bayes and SVM which gives 63% and 75% accuracy, respectively [15].

A significant amount of prior and ongoing studies is based on fake news detection. The misleading information has always been a serious concern worldwide due to its bad influence on social, religious, educational and civilization, and many more fields. We studied research papers in order to carry out this study and filtered out the papers for extensive literature review and summarized them. The proposed model in this study has been drawn from various research being done, and the machine learning algorithms that have been applied in research papers to propose a better way for detection of fake news. We have taken the supervised algorithms which performed best in various study given in [2, 5, 6, 7, 9, 11 and 15] and implemented those algorithms to find out the results.

3 Methodology

Various stages involved in this experiment are given in the following flowchart Fig. 1.

3.1 Dataset

We obtained the dataset for this study from the Kaggle website [16] which contains two files. Out of these two, one contains real news articles and another one contains fake news articles. Real articles are around 21,417 and fake articles are around 23,481

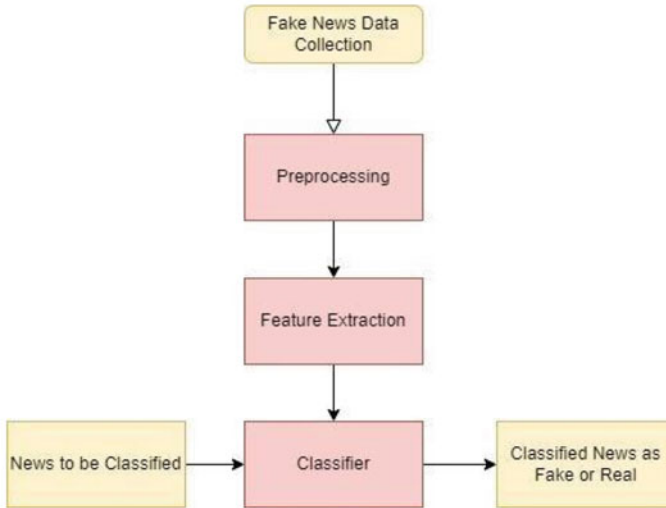


Fig. 1 Flowchart for fake news detection

	title	text	subject	target
0	While You Were Distracted, Republicans Gutted...	On Monday, Missouri Gov. Eric Greitens signed ...	News	fake
1	WATCH WHAT HAPPENS WHEN CHRISTIAN MAN ASKS 13 ...	I was wondering if I could get a cake that sa...	politics	fake
2	LISTEN: Dispatch Audio Reveals Cop Who Killed...	As much as Jeronimo Yanez, the police officer ...	News	fake
3	South Africa's Zuma asks court to set aside re...	JOHANNESBURG (Reuters) - South African Preside...	worldnews	true
4	Bombs kill at least 17 people in Somali capita...	MOGADISHU (Reuters) - Two car bombs killed at ...	worldnews	true

Fig. 2 Head of the data

with a total of 23,481. To further proceed with both the datasets, we had combined the dataset that contains the combination of both fake as well as real. Figure 2 shows the head of our data.

The dataset we are using is having five features, i.e., title, text, subject, date, and target. So, we dropped the unnecessary columns, i.e., date and title as we will be working with only text here.

The data we have used here is available in text as we know that text data requires preprocessing in order to be changed over into a suitable structure for information display. There are several methods for transforming text data, including natural language text processing approaches, which we employed. After removing the date and title columns that were no longer needed for modeling and converting the text to lowercase. We have performed stop word removal because there are many words in a text that occur very frequently in a document and have not much information such as ‘a,’ ‘is,’ ‘the,’ and ‘am.’ To improve the accuracy of analysis, these words are generally ignored using natural language toolkit (NLTK) library for stop word removal. After that, here are many words in a text that occur very frequently in a document

and have no much information such as ‘a,’ ‘is,’ ‘the,’ and ‘am.’ To improve the accuracy of our analysis, these words are ignored using natural language toolkit (NLTK) library for stop word removal. After that, punctuation removal was performed as punctuation like commas and full stops don’t add much importance to text so they should be filtered out. We split the data into train and test before feeding it into the machine learning model. We have separated the 30% data into test set and 70% into train set. A subset of the dataset used to train the model has already revealed the outcome. The detection model is tested on a subset of the dataset, and the test set is utilized to forecast the outcome. For feature extraction, we have concentrated on two distinctive choice techniques: term frequency and term frequency inverted document frequency. This retrieval methodology considers the frequency of a phrase as well as the inverse document frequency.

A term’s frequency in a text may be determined using term frequency. The term n in the formula represents the number of times the phrase appears in each document or text. As a result, each term has a TF value.

3.2 Machine Learning Algorithms

In this paper, we are presenting five different supervised machine learning classification algorithms. The following is a quick rundown of all the algorithms used:

3.2.1 Logistic Regression

It’s a tool for categorizing binary data. For binary classification, usually, linear regression is used to create the best fit line. When two classes can be separated linearly, logistic regression is used. It is within the supervised machine learning algorithm category. It’s a machine learning-based categorization problem-solving approach. In logistic regression, a type of predictive analysis, the probability assumptions are applied. To complete a binary classification job, a linear equation is used as input, and the logistic function and log odds are used in the logistic regression model. It employs a more complicated function when compared to linear regression.

$$y = d_0 + d_1 * x \tag{1}$$

$$Q = 1/(1 + e^{-y}) \tag{2}$$

$$\ln(Q/(1 - Q)) = d_0 + d_1 * x \tag{3}$$

where in Eq. (3), d_0 is slope, d_1 is intercept, and x are a data point. Equation (2) is a sigmoid function where Q has been used to eliminate the outlier’s effect.

3.2.2 Decision Tree

It's an ML supervised classification algorithm which means we have to clarify what the information is and what the relating yield is in the preparation information. It is a tree-like construction where the information is consistently parted by a specific boundary. The elements of a dataset are addressed by the inner nodes and branches that address the decision rules and each leaf addresses the ultimate results or choices.

3.2.3 Naïve Bayes

The Naive Bayes approach, a supervised machine learning methodology based on the well-known Bayes theorem, is used to tackle classification issues. It is most commonly used for text classification with a big training dataset. One of the most simple and effective classification methods is the Naive Bayes classifier. It allows for the rapid building of machine learning models as well as effective training and testing to make speedy predictions. It's a probabilistic classifier, which implies the algorithm's whole basis is built on probabilities that have been computed, and it predicts based on an item's likelihood.

Naïve Bayes Equation:

$$P(R|S) = P(S|G) * P(R)/P(S)$$

where $P(R|S)$ is the posterior probability. $P(S|R)$ is the likelihood. $P(R)$ is the class prior probability. $P(S)$ is the predictor of prior probability.

3.2.4 Random Forest

It is a supervised ML technique. It is basically established on the outfit learning techniques where different classifiers are united to deal with an issue and to chip away at the display of the presentation of the model. Random Forest is a classifier that calculates the dataset's predicted precision by averaging the results of many decision trees applied to different subsets of the dataset.

3.2.5 Support Vector Machine

This approach aims to find a hyperplane (where N is the number of characteristics) that clearly arranges the principal elements in an N -dimensional space. There is an assortment of hyperplanes from which to isolate the two kinds of informative items. Our point is to track down the plane with the biggest edge or distance between relevant items from the two classes. Boosting the edge distance gives some support, making it more straightforward to arrange ensuing information points.

3.3 *Evaluating Measures*

Evaluation metrics are frequently used to assess categorization performance. As a result, performance measurements are the most common. So, we have used different metrics to evaluate our classifiers given as follows.

3.3.1 Accuracy

It gives the comparison of actual and predicted labels, i.e., it measures how often a classifier predicts accurately. It can be formulated as

$$\text{Accuracy} = \text{Correct prediction} / \text{Total data points}$$

3.3.2 Precision

Precision is a proportion to tell how exact a classifier is performing. Precision P can be formulated as the ratio of total true positives to total predicted instances

$$\text{Precision}(P) = \text{TP} / (\text{TP} + \text{FP})$$

3.3.3 Recall

It tells what percentage of positive instances were successfully identified and its formula is

$$\text{Recall}(R) = \text{TP} / (\text{TP} + \text{FN})$$

3.3.4 F-Measure

It is represented as a harmonic mean of precision and recall and can be formulated as

$$F - \text{Measure}(F) = 2P \cdot R / (P + R)$$

where

TP (True Positive) belongs to a class it belongs actually,

FP (False Positive) belongs to a class it doesn't belong actually, FN (False Negative) doesn't belong to a class it actually should belong, and TN (True Negative) doesn't belong to a class it actually doesn't belong.

4 Implementation and Results

In Figs. 3, 4, 5, 6 and 7, we are giving the various confusion matrices that we obtained after applying five supervised machine learning algorithms. As discussed in methodology, each confusion matrix contains the four values TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative). In our experiment, TP represents that the news which was actually fake is also predicted as fake. FP represents that the news which was actually real is predicted as fake. FN represents that the news which was actually fake but is predicted as real. TN represents that the news which was actually real is also predicted as real. The real and fake true label and predicted label are shown in confusion matrix. We have implemented five supervised machine learning algorithms here, i.e., logistic regression (LR), decision tree (DT), Naive Bayes (NB), Random Forest (RF), and support vector machine (SVM).

Figures 8 and 9 represent the word cloud obtained from both real and fake news sets, respectively, to represent the significant textual data points/words in the dataset.

Table 1 indicates the accuracy % and performance metric of all five classifiers, i.e., precision, recall, and F1-score.

As described in the methodology, we have implemented all the five ML classification algorithms and calculated the accuracy and performance measures. From

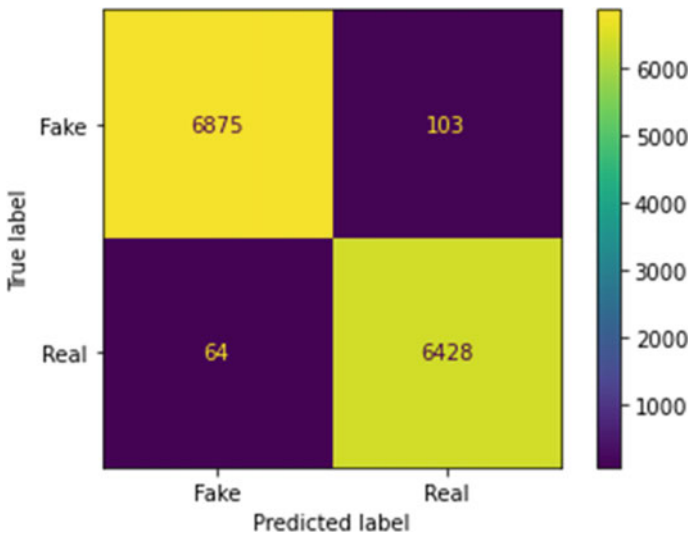


Fig. 3 Confusion matrix using LR

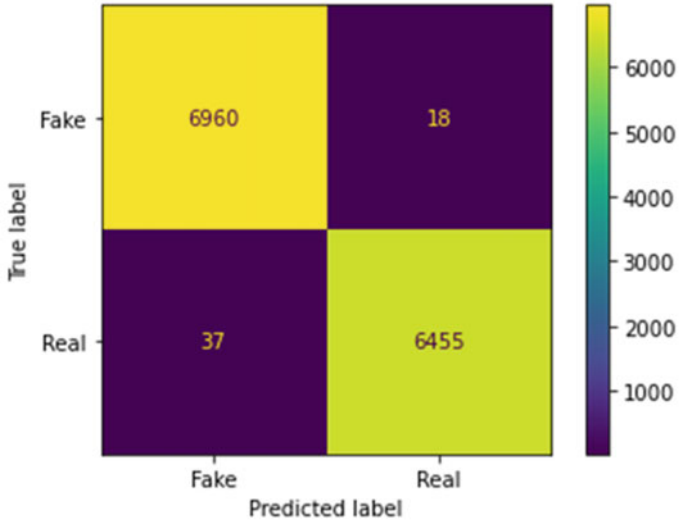


Fig. 4 Confusion matrix using DT

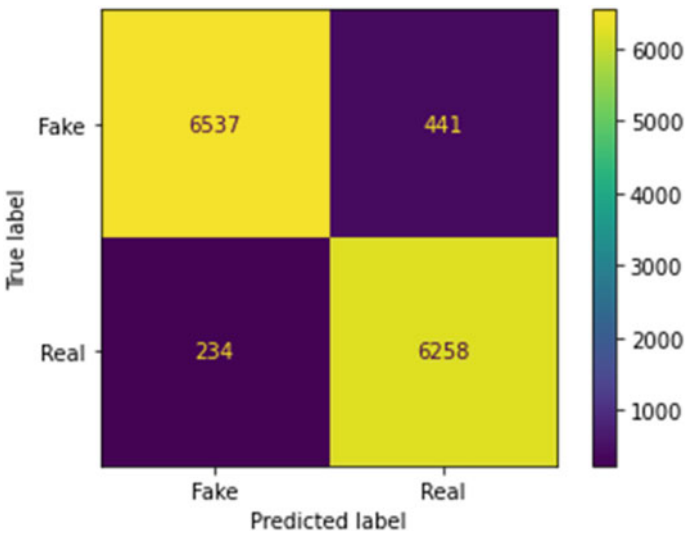


Fig. 5 Confusion matrix using NB

Table 1, we can see that decision tree outperforms here which has the accuracy 99.59% and all performance metrics (precision, recall, and F1-score) as 1.00 perform the best. After decision tree, SVM performs with a very negligible difference in accuracy when compared. Naïve Bayes has the lowest accuracy, i.e., 94.99% and all performance metrics, i.e., (recall, F1-score, and precision) as 0.95.

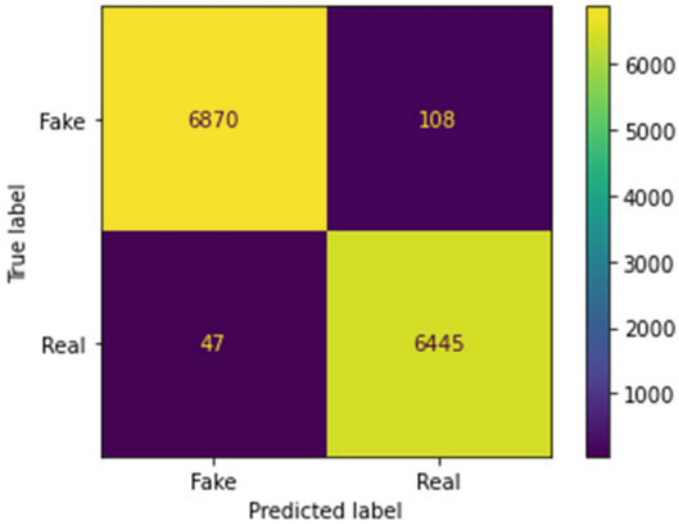


Fig. 6 Confusion matrix using RF

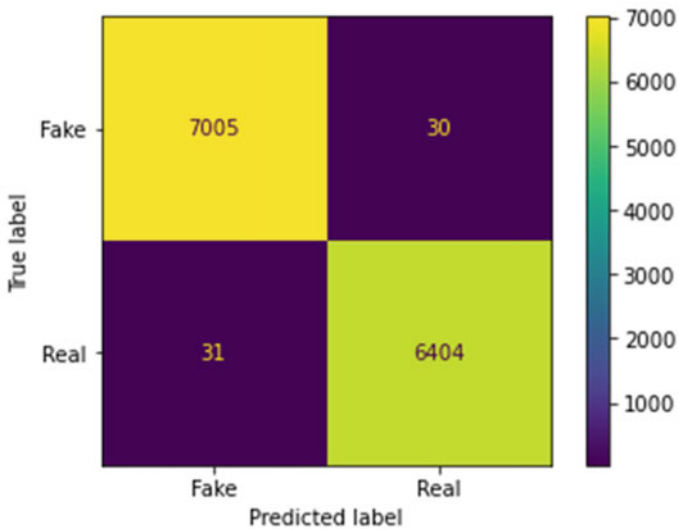


Fig. 7 Confusion matrix using support vector machine

Figure 10 graph shows the comparison of all the models used in this experiment which shows that decision tree has the highest accuracy of 99.59% and the Naïve Bayes classifier has the lowest accuracy of 94.99%.

In our experiment, decision tree is performing best. As the data is categorical that is the news is either fake or real, so in such cases this algorithm performs

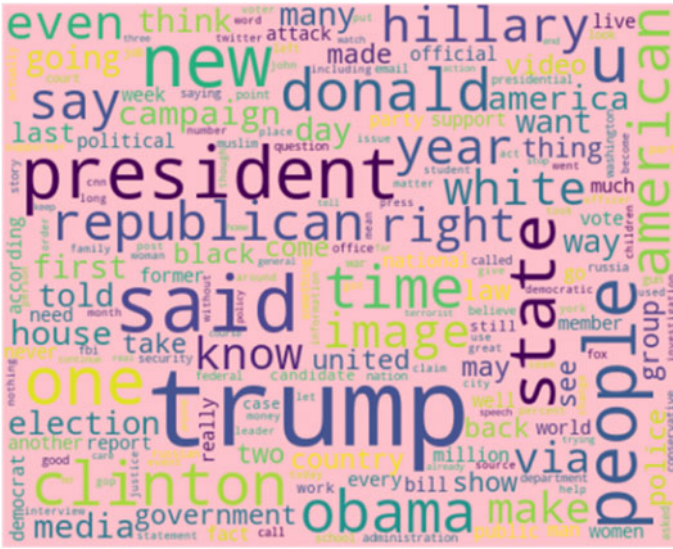


Fig. 8 Word cloud for real news

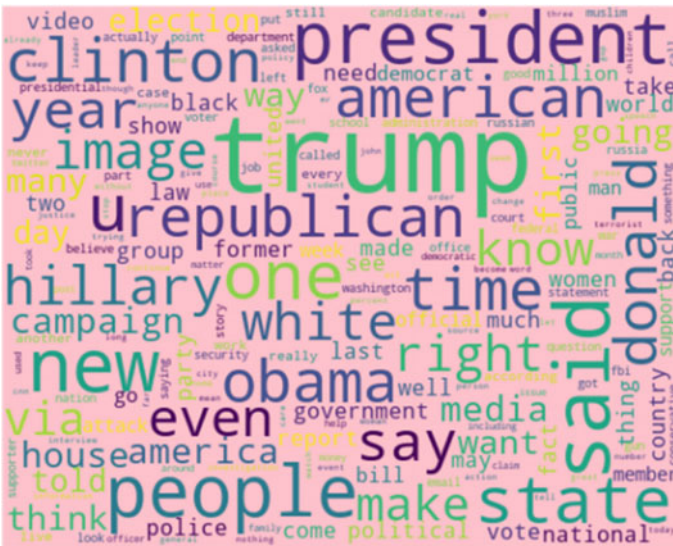


Fig. 9 Word cloud for fake news

Table 1 Accuracy and performance measures

ML classifier	Accuracy (%)	Precision	Recall	F1 Score
Logistic regression	98.76	0.99	0.99	0.99
Decision tree	99.59	1.00	1.00	1.00
Random forest	98.85	0.99	0.99	0.99
Naïve Bayes	94.99	0.95	0.95	0.95
Support vector machine	99.58	1.00	1.00	1.00

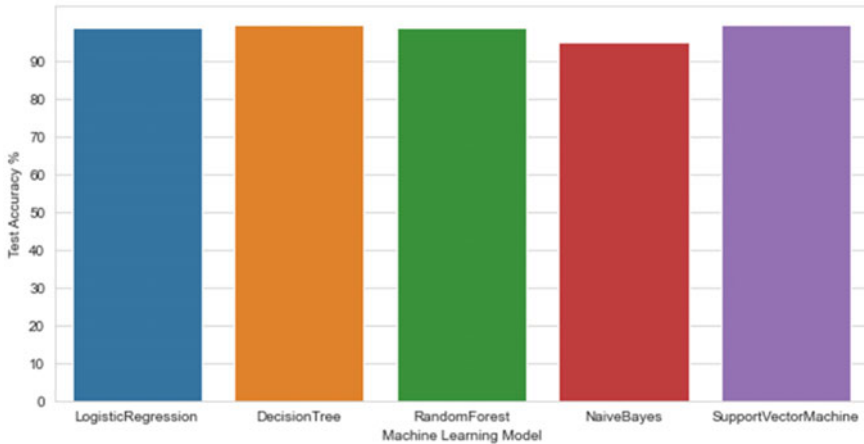


Fig. 10 Plot for comparison of all the classifiers used

better as compared to other supervised ML algorithm. When compared, NB is a generative model while DT is discriminative model. When compared, SVM solves nonlinear issues using the kernels method, whereas decision trees handle the problem by deriving hyper-rectangles in input space.

Hence in our study, decision tree is performing best with an accuracy of 99.58%.

5 Conclusion and Future Scope

Fake news has a huge influence on our social life, as well as in other domains, such as politics and education. Fake news may create significant social and societal harm, as well as have potentially dangerous consequences. It is becoming more difficult for the citizens/consumers to obtain the information that is precise and error free and reliable because of increasing the dimensions of social media. It’s critical to discover such false information early on in order to avoid the global harm it can do. As a result, in this paper, we designed a methodology for detecting false news that combines NLP

techniques with supervised learning classification algorithm. In this work, we have presented a machine learning approach using various machine learning classifiers to detect fake news. After comparing the performance of each model, the conclusion can be drawn that decision tree outperforms the other algorithms being used, i.e., with the accuracy of 99.59%, and secondly, SVM performs well with the accuracy of 99.68%. This approach would be helpful to identify fake news effectively and with higher accuracy in future.

Future work could include comparing multiple deep learning approaches and new ensemble learning methods to the classification techniques used in this study and determining the best strategy for detecting fake news. Also, we may integrate a larger dataset from different sources like various URLs and news publication sites as it would be having bigger journalese and could be used for obtaining better results in a generalized manner.

References

1. Granik M, Mesyura V (2017) Fake news detection using naive Bayes classifier. In: 2017 IEEE 1st Ukraine conference on electrical and computer engineering UKRCON, pp 900–903. <https://doi.org/10.1109/UKRCON.2017.8100379>
2. Ahmed H, Traore I, Saad S (2017) Detection of online fake news using N-Gram analysis and machine learning techniques. In: Traore I, Woungang I, Awad A (eds) Intelligent, secure, and dependable systems in distributed and cloud environments, pp 127–138. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-69155-8_9
3. Campan A, Cuzzocrea A, Truta TM (2017) Fighting fake news spread in online social networks: actual trends and future research directions. Presented at the December 1. <https://doi.org/10.1109/BigData.2017.8258484>
4. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2018) Automatic detection of fake news. COLING 2018 - 27th Int. Conf. Comput. Linguist. Proc. 3391–3401 (2018). <https://doi.org/10.48550/arXiv.1708.07104>
5. Aphiwongsophon S, Chongstitvatana P (2018) Detecting fake news with machine learning method. In: 2018 15th International conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON), pp 528–531. IEEE, Chiang Rai, Thailand. <https://doi.org/10.1109/ECTICon.2018.8620051>.
6. Gahirwal M (2008) Fake News Detection 3
7. Ozbay FA, Alatas B (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Phys A Stat Mech Appl 540:123174. <https://doi.org/10.1016/j.physa.2019.123174>
8. Agarwal V, Sultana HP, Malhotra S, Sarkar A (2019) Analysis of classifiers for fake news detection. Proc Comput Sci 165:377–383. <https://doi.org/10.1016/j.procs.2020.01.035>
9. Reis JCS, Correia A, Murai F, Veloso A, Benevenuto F (2019) Supervised learning for fake news detection. IEEE Intell Syst 34:76–81. <https://doi.org/10.1109/MIS.2019.2899143>
10. Zhang X, Ghorbani AA (2020) An overview of online fake news: characterization, detection, and discussion. Inf Process Manage 57:102025. <https://doi.org/10.1016/j.ipm.2019.03.004>
11. Shaikh J, Patil R (2020) Fake news detection using machine learning. In: 2020 IEEE international symposium on sustainable energy, signal processing and cyber security (IS-SSC), pp 1–5. IEEE, Gunupur Odisha, India (2020). <https://doi.org/10.1109/ISSC50941.2020.9358890>

12. Smitha N, Bharath R (2020) Performance comparison of machine learning classifiers for fake news detection. In: 2020 Second international conference on inventive research in computing applications (ICIRCA), pp 696–700. IEEE, Coimbatore, India (2020). <https://doi.org/10.1109/ICIRCA48905.2020.9183072>
13. Kesarwani A, Chauhan SS, Nair AR (2020) Fake news detection on social media using K-nearest neighbor classifier. In: 2020 International conference on advances in computing and communication engineering (ICACCE), pp 1–4. IEEE, Las Vegas, NV, USA. <https://doi.org/10.1109/ICACCE49060.2020.9154997>
14. Khanam Z, Alwasel BN, Sirafi H, Rashid M (2021) Fake news detection using machine learning approaches. In: IOP conference series: materials science and engineering, vol 1099, p 012040. <https://doi.org/10.1088/1757-899X/1099/1/012040>
15. Nagaraja A, KN S, Sinha A, Rajendra Kumar JV, Nayak P (2021) Fake news detection using machine learning methods. In: International conference on data science, E-learning and information systems 2021, pp 185–192. ACM, Ma'an Jordan. <https://doi.org/10.1145/3460620.3460753>
16. <https://www.kaggle.com/clmentbisailon/fake-and-real-news-dataset>