Machine learning for fake news classification with optimal feature selection

Muhammad Fayaz1 · Atif Khan2 · Muhammad Bilal2 · Sana Ullah Khan3

Accepted: 10 January 2022 / Published online: 29 January 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Nowadays, current events related to diverse fields are published in newspapers, shared on social media and broadcasted on radio and television. The explosive growth in online news contents has made it very difficult to discriminate between real and fake. As a result, fake news has become prevalent and immensely challenging to analyze and verify. Indeed, it is a big challenge to the government and public to debate the situation depending on case to case. For this purpose, a mechanism has to be taken on fact-checking rumors and statements particularly those that get thousands of views and likes before being debunked and refuted by expert sources. Various machine learning techniques have been used to detect and classify fake news. However, these approaches are restricted in terms of accuracy. This study has applied a random forest (RF) classifier to predict fake or real news. For this purpose, twenty-three (23) textual features are extracted from ISOT Fake News Dataset. Four best feature selection techniques like chi2, univariate, information gain and feature importance are used to select fourteen best features out of twenty-three. The proposed model and other benchmark techniques are evaluated on benchmark dataset using best features. Experimental findings show that the proposed model outperformed state-of-the-art machine learning techniques such as GBM, XGBoost and Ada Boost Regression Model in terms of classification accuracy.

Keywords Machine learning · Random forest · Fake news · Feature selection

1 Introduction

Internet offers many possibilities with many challenges regarding news. The number of communication channels is growing over time. In addition to conventional channels such as newspapers and TV, news transmission through blogs and social networks have been arisen via Internet. It has become simpler for customers to receive the latest news on their fingertips through social sites. These platforms are useful to share ideas and discuss diverse issues of governance, education and health. Many organizations use various sites for monetary purposes too. These sites are also highly effective for news transmission. But with the absence of editorial board like that of newspapers, fake news are propagated on these sites that has become a great challenge. A study conducted by Twitter reveals that fake news is 100 times quickly spread and shared than real news. Fake news affect millions of people in certain countries (Vogel and Meghana 2020). Such propaganda and rumors news affects stock prices, stock purchases, investment plans and even reactions to natural disasters. The contents of fake news are framed in such a way that it may create public opinions and fully win over the reader to make them completely confused and divert their attention from real news (Hakak et al. 2021). Detection of fake news is a challenging task as it requires rationalism. Many fact-checking websites are deployed to reveal fake news to counter the growing misinformation. Such websites play a critical role in clarifying false news but very time-consuming and need expertise. Therefore, it is quite challenging to detect and analyze the data authenticity (Napoli 2018).
This study uses benchmark and other machine learning approaches for fake news classification. This research uses random forest (RF) classifier to improve the accuracy of fake news classification. The proposed machine learning model operates as follows: first, we extract twenty-three (23) textual features from the ISOT Fake News Dataset and the news dataset is represented as a features vector. Too many features influence the model efficiency and performance, and not all features have the same predictive model contribution. So it is important to strip out non-valuable and less important features to increase model accuracy and reduce its complexity. Therefore, four different feature selection techniques are used, such as chi^2, feature importance, information gain and univariate to pick the fourteen (14) best features out of 23 features. Based on best features, the efficiency of the proposed model is compared with the benchmark techniques.

This study contributes in the following way:

- To apply a random forest (RF) classifier for classification of news as fake and real.
- To evaluate the efficacy of RF classification model and other benchmark ML models with all textual features (twenty-three) extracted from ISOT fake news dataset.
- To evaluate the efficacy of the proposed RF model and other benchmark ML models with respect to the best features extracted using four different feature selection techniques (chi^2, univariate, feature importance, information gain).

Remaining sections of the paper are structured as follows: Sect. 2 is related to comprehensive literature study. Section 3 describes the proposed machine learning model. Section 4 is related to various experimental findings and complete discussion. Conclusion and future recommendations are given in Sect. 5.

2 Related work

News are very important because it keeps the public informed about activities and events around and beyond their premises. Reports showed that most adults use digital forms such as social media and web/search engines to access news instead of traditional media. Fake news detection has got considerable attention (De Choudhury et al. 2014). Numerous methods have been suggested to detect fake news using various types of features and datasets. Authors in (Okoro et al. 2018) proposed a Machine-Human (MH) model for detecting fake news in social media by combining a human literacy news detection tool with machine linguistic and network-based approaches. The authors (Khan et al. 2021) used CNN, LSTM, Bi-LSTM, C-LSTM, Heterogeneous Graph Neural Network (HAN), Cov-HAS, Char-level C-LSTM models to compare the Glove and character embeddings for detection of fictitious news on three datasets. They also used n-gram features with naive bayes classifier for detection of fabricated news. It was discovered that n-gram features combined with Naive Bayes classifier yield very promising results for detecting fabricated news, which are nearly equivalent to the performances of CNN-based models.

The authors in (Ozbay and Alatas 2020) compared supervised machine learning algorithms for detecting fake news in online social media. Three different real-world datasets were used to evaluate the algorithms. The Decision Tree Algorithm outperformed the other algorithms in terms of accuracy, precision, and F-measure. The authors of (Gravanis et al. 2019) tested several classification models with an enhanced set of linguistic features on five different datasets containing both fake and real news. According to the empirical results, Adaboost achieved 95% accuracy across all datasets, while support vector machine (SVM) and Bagging algorithms came in second and third place, respectively.

The study (Ahmad et al. 2020) used various machine learning and ensemble techniques to detect fake and real news from various domains. The results showed that the ensemble model XGBoost outperformed other classifiers in terms of accuracy. The authors of (Ahmed et al. 2019) investigated various machine learning algorithms for detecting fake news in social networks. The authors in (Ruchansky et al. 2017) proposed the CSI model, which is made up of three modules: Capture, Score, and Integrate. The first module, which is based on the response and text, employs a recurrent neural network (RNN) to capture the temporal pattern of user activity on a specific article. The second module learns the source characteristic based on users behaviors, and the two are combined with the third module to determine whether an article is fake or not. Experimentation on real-world data shows that CSI outperformed the existing models in terms of accuracy and extracted meaningful latent representations of both users and articles. The goal of this study (Mansouri et al. 2020) is to detect fake news using deep learning techniques. First, various features of text and image data are extracted using CNN in this method. Then, linear discrimination analysis (LDA) is employed to predict the classes of unclassified data. The proposed method outperformed other methods in terms of recall, precision, sensitivity and specificity.

This study (Najar et al. 2019) used finite mixture models of Dirichlet Compound Multinomial (EDCM) distributions to tackle the problem of detecting fake news. For the learning of these mixture models, they developed a Bayesian approach based on Metropolis–Hastings algorithm and Markov Chain Monte Carlo. The authors in (Jain et al.
2019; Reis et al. 2019) extracted different textual features like language features such as n-gram and part of speech tagging, lexical features (character and word-level signals), psycholinguistic features, semantic features, subjectivity and sentiment scores extracted from fake news dataset. The features vectors are then fed to different classification algorithms for detection of fake news. Results showed that random forest and XGB performed superior than other classifiers.

The authors in (Faustini and Covões 2019) proposed using One-Class Classification (OCC) to detect fake news by training a model with only fake samples in the training dataset. We compared a novel algorithm called DCDistanceOCC to others published in the literature and found similar results. In a study (Hlaing and Kham 2020), multidimensional fake news (news content, social engagement, and news stance) was detected using synonym-based features and three different classifiers: Decision Tree classifiers, AdaBoost classifiers, and random forest classifiers. The experimental results show that random forest outperformed the other two classifiers on the social media dataset. According to the study (Mahir et al. 2019), SVM outperformed other classifiers such as Naive Bayes, RNN/LSTM, and Logistic Regression in recognizing fake news extracted from Twitter. The authors in (Al-Ash et al. 2019) employed an ensemble learning approach to distinguish between Indonesian fake news and real news, as well as to address the imbalanced data issue that encountered on the given dataset. The results showed that the random forest classifier outperformed the multinomial naive bayes and the support vector machine. The study of (Katsaros et al. 2019) eight different ML and DL models were evaluated for classification of fake news. The result showed that the CNN is superior than other models.

The study of (Choudhary et al. 2021) introduced BerConvoNet, a deep learning framework for classifying given news text as real or fake with negligible error. The framework consists of two main building blocks: a news embedding block (NEB) and a multi-scale feature block (MSFB). NEB extracts word embeddings from news articles using Bidirectional Encoder Representations from Transformers (BERT). Following that, these embeddings are fed into MSFB, which extracts various features from news word embedding. The MSFB output is fed into a fully connected layer for classification. The results demonstrated that BerConvoNet outperformed other models on various performance metrics. The authors in (Vogel and Meghana 2020) extracted various hand-crafted features from news datasets and used various classification models to classify fake news. According to the results, SVM achieved the highest accuracy of 92% in the classification of fake news.

## 3 Proposed methodology

This section presents the architecture of proposed model for classification of fake news as shown in Fig. 1. It consists of three main phases: first phase of preprocessing comprises of Sentence segmentation, tokenization, stopword removal and word stemming, the second phase includes features extraction and best features selection using well-known feature selection techniques (chi², feature importance, information gain and univariate) and the final phase illustrates the classification of fake news using proposed RF model and other machine learning models.

### 3.1 Preprocessing

To improve the model’s performance, the data are preprocessed/cleaned before being fed to it. This phases involves various steps which are briefly discussed as follows.

#### 3.1.1 Sentence Segmentation

This step identifies the text borders and breaks the text into sentences. Exclamation (!), interrogation (?), and utter stop (.) signs are widely used as markers to segment the paragraph/document into sentences.

#### 3.1.2 Tokenization

In this step, the phrases and sentences are divided into separate tokens (words). Tabs, blanks, and punctuation symbols such as dot (.), comma (,), semicolon (;), colon (:), are used as main indicators for splitting sentences into tokens.

#### 3.1.3 Stopwords removal

Stopwords are frequently occurring words with little meaning in the text documents. Prepositions (in, on, at, etc.), conjunctions (and, thus, too, etc.), and articles (a, an, a) are examples of stopwords that are typically removed from the text to reduce computational overhead and improve system performance.

#### 3.1.4 Word stemming

Word stemming plays a significant role in preprocessing. In order to normalize the word/token to a standard form, this step changes the derived words to its base or stem word. The famous stemming algorithm, Porter’s stemming, is adopted to remove suffixes like -ing, -es, -ers from the text.
words. For example, the words ‘looking’ and ‘looks’ will be changed to its base form ‘look’ after stemming.

3.2 Features extraction

Text cannot be understood directly by ML algorithms and must be encoded in a format that ML algorithms can use. The numeric vectors can be understood by ML algorithms. As a result, the textual data must be transformed into real-valued vectors, and this process is known as feature extraction. The real-valued vectors are then fed into machine learning models. In this study, we extracted twenty-three (23) features form ISOT fake news dataset. Almost all the extracted features are textual and are shown in Table 1.

3.2.1 Features selection

It is normally not a good idea to use all twenty-three (23) features to classify the ISOT News dataset as fake or real News because extracting a large number of features is computationally expensive. When developing a consistent and effective statistical model, not all features have the same significance and weight. Some features are more useful and contribute more to model prediction and play a critical role in classification accuracy, whereas others are less valuable and have the least impact on model performance. In addition, the appropriate and useful features eliminate over-fitting, increase precision and reduce the predictive model training time. In this study, we used four feature selection techniques on the ISOT News dataset, including chi^2, univariate, information gain, and feature importance, to extract the fourteen best features, as shown in Table 2.

3.3 Classification models

This section aims to classify ISOT news as fake and real news using random forest and other ML Models. We trained and evaluated the classifiers on the ISOT Fake News dataset using tenfold cross-validations in order to validate the impact of the individual models and the proposed model in the context of all textual features and best features selected by features selection techniques. Random forest is a traditional machine learning algorithm that is used for classification and regression problems by ensembling a large number of decision trees. It predicts model accuracy using bagging and bootstrap methods. Each decision tree’s prediction is combined for a final prediction using a majority of votes as shown in Fig. 2.

4 Experimental settings

The proposed machine learning model is tested on the ISOT News Dataset, which contains a total of 44,919 news items, 23,502 of which are fake news and the remaining 2147 are real news. The dataset is preprocessed by splitting the news text into sentences. The sentences are tokenized into words and stopwords are removed. Initially, twenty-
### Table 1 List of all features extracted from ISOT News dataset

| S. no | Features | Description |
|-------|----------|-------------|
| 1     | Counts words | Count total numbers of words |
| 2     | Uppercase character | Count total number of uppercase characters |
| 3     | Lowercase character | Count total number of lowercase character |
| 4     | Character count | Count total number of characters |
| 5     | Count sentence | Count total number of sentences in fake News |
| 6     | Automated readability index | Automated readability index is readability test |
| 7     | Coleman–Liau index | Coleman–Liau index is text readability test |
| 8     | Flesch reading ease | Flesch reading ease is text readability test |
| 9     | Flesch-Kincaid grade | Flesch-Kincaid grade level is text readability test |
| 10    | Dale–Chall formula | Dale–Chall formula is text readability test |
| 11    | Topic | Topic related words |
| 12    | Count spaces | Count number of spaces |
| 13    | Part of speech | Count numbers of part of speech |
| 14    | Principal component analysis | Principal component analysis |
| 15    | Gunning fox index | Gunning fog index is readability test |
| 16    | SMOG formula | Mc-Laughlin’s SMOG formula is readability test |
| 17    | Sentiment analysis | Sentiment analysis show text positivity and negativity |
| 18    | Bag-of-word | Bag of words is a text representation representing the presence of words |
| 19    | Tf-idf (unigram) | TF-IDF with unigram |
| 20    | Tf-idf (bi-gram) | TF-IDF with Bi-GRAM |
| 21    | Tf-idf (tri-gram) | TF-IDF with Tri-Gram |
| 22    | Stopwords | Count total numbers of stop words |
| 23    | Negative words | Count total numbers of negative word |

### Table 2 List of fourteen best features chosen by different features selection techniques for ISOT News dataset

| S. no | Features selected by chi² | Features selected by Univariate | Feature importance | Information gain |
|-------|---------------------------|--------------------------------|--------------------|------------------|
| 1     | Counts words | Counts words | Counts words | Count unique words |
| 2     | Uppercase character | Uppercase character | Uppercase character | Uppercase character |
| 3     | Lowercase character | Lowercase character | Count sentence | Count sentence |
| 4     | Character count | Character count | Flesch-Kincaid grade level | Flesch-Kincaid grade level |
| 5     | Count sentence | Count sentence | Flesch reading ease | Flesch reading ease |
| 6     | Automated readability index | Automated readability index | Dale–Chall formula | Dale–Chall formula |
| 7     | Coleman–Liau index | Coleman–Liau index | Gunning fog index | Gunning fog index is readability |
| 8     | Flesch-Kincaid grade level | Flesch-Kincaid grade level | Songs-formula | SMOG formula |
| 9     | Flesch reading ease | Dale–Chall formula | Topic | Count lowercase characters |
| 10    | Principal component analysis | Topic | Tf-idf (unigram) | Tf-idf (unigram) |
| 11    | Tf-idf (unigram) | Count-spaces | Stopwords | Tf-idf(bi-gram) |
| 12    | Tf-idf (bi-gram) | Part of Speech | Negative words | Tf-idf(tri-gram) |
| 13    | Tf-idf (tri-gram) | Tf-idf (unigram) | Part of Speech | Stopwords |
| 14    | Dale–Chall formula | Sentiment analysis | Sentiment analysis | Negative words |
three (23) textual features are selected from the ISOT News dataset for fake news classification. The proposed model and other machine learning models are evaluated on all twenty-three features. All features are not equally important in developing a consistent and accurate predictive model. Some features are more important and contribute more to model accuracy, while others are less important and have a negative impact on model performance. Furthermore, appropriate and useful features reduce over-fitting, increase precision, and shorten predictive model training time. Therefore, we used four features selection technique such as chi², univariate, information gain and feature importance, to reduce the feature space and choose the fourteen best features. Finally, on the benchmark dataset, the proposed model and other benchmark techniques are evaluated on all features and best features in terms of classification accuracy.

5 Results and discussion

The performance of the random forest (RF) model and other machine learning (ML) models is evaluated in the first experiment using all twenty-three (23) textual features extracted from the ISOT News dataset. Table 3 shows the results of this experiment.

The results show that the proposed RF model has the highest score of 97.25% on all features when compared to other classifiers.

Then, we ran the second experiment, and the top 14 best features were chosen using the chi² features selection technique. All models, including the proposed model, were evaluated using the best features available, and the results are shown in Table 4. The proposed model achieved 97.33% accuracy and outperformed individual models on fourteen (14) best features for the task of fake news classification on the ISOT dataset. However, the logistic regression got the lowest accuracy of 45.54%.

The third experiment was then run, and the top 14 best features were chosen using the univariate feature selection technique. All models, including the proposed model, were
evaluated using the 14 best features, with the results shown in Table 5.

The results show that the proposed model outperformed individual models, with an accuracy score of 97.27% on best features chosen using the univariate features selection technique and Naïve Bayes Gaussian achieved the lowest accuracy of 43.48%.

The fourth experiment was then run, and the fourteen (14) best features were chosen using the feature importance technique. Table 6 shows the results of an evaluation of all models, including the proposed model, on the top 14 best features. The results show that the proposed model outperformed other classifiers. The proposed model’s accuracy is 96.60%.

The fifth experiment was then carried out, and the fourteen best features were chosen using the information gain method of feature selection. All models, including our proposed model, are evaluated using the top 14 best features from the ISOT dataset, and the results are recorded in Table 7. The proposed model had the highest accuracy of 96.42%, while Naive Bayes had the lowest.

From experimental results shown in Fig. 3 and Table 8, the following conclusion are drawn:

- The accuracy of the proposed model (random forest) is improved with best features selected using chi2 and univariate. However, the when features are selected using feature importance and information gain features selection techniques, the accuracy of the proposed model is slightly reduced.
- The accuracy of boosting technique like XGBoost, GBM and Ada Boost did not improve on best features selected using different features selection techniques.
Performance of Gaussian Naïve Bayes is significantly improved on best features selected using chi2 as compared to other features selection techniques.

Overall, in terms of classification accuracy, the proposed model outperformed all individual models, as well as other boosting approaches.

6 Conclusion and future work

Fake news detection is a challenging task in the area of text classification. Many attempts have been made by various researchers to address this issue. This study proposed random forest as machine learning classifier to classify the news as fake or real. In this connection, twenty-three (23) features were extracted from the ISOT fake news dataset.
In order to reduce the feature space and filter out the best features, we used four feature selection techniques like chi\(^2\), univariate, features importance and information gain, to choose the fourteen (14) best features. The proposed RF model and other ML models were evaluated in the context of fake news classification task using all features and fourteen (14) best features. The experimental results show that proposed model outperformed all other classifiers in terms of better classification accuracy. In future, we intend to employ deep ensemble models for fake news classification.

**Acknowledgements** The authors did not receive support from any organization for this work.

**Declarations**

**Conflict of interest** The authors declare that they have no conflict of interest.

**References**

Ahmad I, Yousaf M, Yousaf S, Ahmad MO (2020) Fake news detection using machine learning ensemble methods. Complexity 2020:1–11

Ahmed S, Hinkelmann K, Corradini F (2019) Combining machine learning with knowledge engineering to detect fake news in social networks-a survey. In: Proceedings of the AAAI 2019 spring symposium, p 8

Al-Ash HS, Putri MF, Mursanto P, Bustamam A (2019) Ensemble learning approach on indonesian fake news classification. In: 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS). IEEE, pp 1–6

Choudhary M, Chouhan SS, Pilli ES, Vipparthi SK (2021) BerConvoNet: a deep learning framework for fake news classification. Appl Soft Comput 110:107614

De Choudhury M, Morris MR, White RW (2014) Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI conference on human factors in computing system, pp 1365–1376

Faustini P, Covões T (2019) Fake news detection using one-class classification. In: 2019 8th Brazilian conference on intelligent systems (BRACIS). IEEE, pp 592–597

Gravanis G, Vakali A, Diamantaras K, Karadasis P (2019) Behind the cues: a benchmarking study for fake news detection. Expert Syst Appl 128:201–213

Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PKR, Khan WZ (2021) An ensemble machine learning approach through effective feature extraction to classify fake news. Futur Gener Comput Syst 117:47–58

Hlaing MMM, Kham NSM Defining news authenticity on social media using machine learning approach. In: 2020 IEEE conference on computer applications (ICCA). IEEE, pp 1–6

Jain A, Shakya A, Khatter H, Gupta AK A smart system for fake news detection using machine learning. In: 2019 international conference on issues and challenges in intelligent computing techniques (ICICT). IEEE, pp 1–4

Katsaros D, Stavropoulos G, Papakostas D Which machine learning paradigm for fake news detection? In: 2019 IEEE/WIC/ACM international conference on web intelligence (WI). IEEE, pp 383–387

Khan JY, Khondaker MTI, Afroz S, Uddin G, Iqbal A (2021) A benchmark study of machine learning models for online fake news detection. Mach Learn Appl 4:100032

Mahir EM, Akhter S, Huq MR Detecting fake news using machine learning and deep learning algorithms. In: 2019 7th international conference on smart computing and communications (ICSCC). IEEE, pp 1–5

Mansouri R, Naderan-Tahan M, Rashti MJ A semi-supervised learning method for fake news detection in social media. In: 2020 28th Iranian conference on electrical engineering (ICEEE). IEEE, pp 1–5

Najar F, Zamzami N, Bouguila N Fake news detection using Bayesian inference. In: 2019 IEEE 20th international conference on information reuse and integration for data science (IRI). IEEE, pp 389–394

Napoli PM (2018) What if more speech is no longer the solution: first amendment theory meets fake news and the filter bubble. Fed Comm LJ 70:55

Okoro E, Abara B, Umagba A, Ajonye A, Isa Z (2018) A hybrid approach to fake news detection on social media Nigerian. J Technol 37:454–462

Ozbay FA, Alatas B (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Phys a: Stat Mech Appl 540:123174

Reis JC, Correia A, Murai F, Veloso A, Benevenuto F (2019) Supervised learning for fake news detection. IEEE Intell Syst 34:76–81

Ruchansky N, Seo S, Liu Y (2017) Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management, pp 797–806

Vogel I, Mehmana M (2020) Detecting fake news spreaders on twitter from a multilingual perspective. In: 2020 IEEE 7th international conference on data science and advanced analytics (DSAA). IEEE, pp 599–606

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.