Multimedia Tools and Applications (2024) 83:30095–30112
https://doi.org/10.1007/s11042-023-16669-z

A fusion of BERT, machine learning and manual approach for fake news detection

Mohammed A. Al Ghamdi[1] · Muhammad Shahid Bhatti[2] · Atif Saeed[2] · Zeeshan Gillani[2] · Sultan H. Almotiri[1]

Received: 27 July 2022 / Revised: 4 July 2023 / Accepted: 23 August 2023 /
Published online: 15 September 2023

Abstract
A large number of users around the globe have preferred to read news and the latest information from the Internet, especially social media, leaving behind the traditional approach of print media. On the one hand, the Internet is a constructive medium to spread the latest news and information briefly. On the other hand, malicious users are very active on the Internet and spread fake news, which becomes viral within a few minutes. The spread of fake news has become a serious threat as many users now rely on Internet news without verification. In this digital world, it is easy to spread any toxic information over the Internet, like hate speech, extremism, propaganda, and political agendas. It is a big challenge in today's digital world to mitigate the spread of fake news; hence, there is a need for an automatic computational tool that can assist in measuring the credibility of news. This study aims to deliver a solution where fake news from Twitter and website-based articles can be detected using the Natural Language Processing (NLP) technique, Bidirectional Encoder Representations from Transformers (BERT), other machine learning classification algorithms, and manual program-based approaches. A dataset with fake and real labels for the textual content is used. Different classification algorithms are evaluated to find a suitable algorithm for delivering a fake news detector. The evaluations are based on machine learning and a program-based approach. The textual content that the user provides, such as an article or tweet, can confirm the legitimacy of fake news. This website offers fake news detection for both website-based news articles and tweets from Twitter in English, Arabic, and Urdu.

Keywords Fake news · BERT · Machine Learning · NLP · Tweets · Articles · Manual Checking

Mohammed A. Al Ghamdi, Muhammad Shahid Bhatti, Atif Saeed, Zeeshan Gillani and Sultan H. Almotiri contributed equally to this work.

✉ Muhammad Shahid Bhatti
msbhatti@cuilahore.edu.pk

Extended author information available on the last page of the article

1 Introduction

With the advancement of technology, online platforms and social media have emerged as new sources of news information for the world. Indeed, it is a benchmark for humanity, as now people can be briefed about anything and anywhere in the world with just a click. But with this vast advancement, there is a catch. As millions of news updates are posted daily, there is no proper mechanism for checking their accuracy. According to an estimation, there are more than 500 million blogs on the internet [1]. Due to its ease and vast availability, people prefer to gather news information from online sources such as website-based articles and social media platforms. These sources have no accountability, and people are using them for their information with a blind eye.

Recently, the world has seen widespread fake news from blogging websites and social media platforms during many significant events. For example, in the recent occurrence of COVID-19, when the whole world was startled by the widespread disease, it was seen that the internet had become a significant source for spreading hazardous fake news such as bogus treatments, vaccine hoaxes, etc. [2]. Other than that, election experts emphasized the issue of fake news in the 2016 U.S. presidential election. For instance, at that time, it was seen that 25% of election-related news stories were incredibly fake and were spreading a specific narrative [3]. Another example is the findings from DisInfoLab. According to DisInfoLab, 265 pro-Indian websites were majorly operational in Europe to spread fake and misleading news and distort the image of Pakistan internationally [4].

Fake news can be defined as any news story that is fabricated and false with no sources, facts, or quotes that can be verified [5]. According to research from Benedictine University, fake news can be categorized as (i) Fake news can be a shared website-based article with distorted headlines on social media platforms. The main purpose of fake news is to generate profits or social presence. (ii) Fake news can be found on websites with unreliable information or misleading content. (iii) websites with click-baiting titles are included for fake news. (iv) Fake news can be in the form of satire or comedy websites [6].

Many researchers have already started to contribute to fighting against fake news by manually verifying the news sources. PolitiFact and Alt News, focused on the United States of America and India, respectively, are considered to be doing a significant job. But it is not possible to manually verify the millions of stories spreading each day.

Researchers and scientists believe that the prevailing problem of fake news can be solved with the help of artificial intelligence and machine learning, as both of these techniques have shown outstanding results on many classification problems, such as detecting spam emails, image recognition, etc.

Many pieces of research have been carried out to solve the problem of fake news on the internet and social media, and each has its findings. Many machine learning algorithms and classifiers have been used for this purpose. In the same way, we have also tried to implement our knowledge of machine learning and data science to fight against this problem using machine learning algorithms and natural language processing techniques. In this way, the project "Fact Check Lab" focuses on contributing to efforts against the widespread dissemination of misleading information and propaganda news. This project will be useful for people who read the news in English and Urdu. The project "Fact Check Lab" is focused on website-based news articles and tweets on Twitter. A website will be created for this project that will be able to take user input in the form of textual news content, either from website-based articles or tweets, and detect whether the news is real or fake.

Research Problem The research's objective is to develop a method for helping users avoid becoming addicted to clickbait by identifying and removing websites that publish fake news. Finding such solutions is essential since they will be beneficial to readers and the IT businesses that are mainly involved in the problem.

Research Objective The objective of this study is to extensively survey different machine learning techniques to properly detect fake news in English, Arabic, and Urdu. Moreover, we are also using the Bidirectional Encoder Representations from Transformers (BERT) model along with a manual approach that is based on the validation of URLs, checking grammar, and checking swear words. The main aim is to contribute to introducing countermeasures against the spread of fake news, which can be in the form of online blogs, articles, or tweets. The objective is to introduce a system that would be able to detect fake news from the textual content copied from online blogs, articles, or tweets from Twitter. The main objectives of the study are given below:

- Evaluate different machine learning algorithms and choose the best-performing algorithm for the training model.
- Training model on BERT.
- Manual-based approach based on validation of URL, checking grammar, and checking the swear words.
- Delivering the detection results based on the percentage calculated by Machine learning, BERT, and manual-based approach. **Proposed Solution** The proposed solution to the issue concerned with fake news entails the use of a tool that can identify fake sites from the results provided to a user by a search engine or a social media news feed. To deal with fake news, a new approach has been proposed which is a fusion of machine learning and manual programmed approach. In machine learning, the system will detect whether the news is fake or not using a trained model of BERT. Other than BERT, a best-performing classifier will also be used. On the other hand, when it comes to manual programmed-based detection, the system checks three parameters. First of all, the domain name is checked. This will check if a domain name is of any highly credible news agency or not. For example, if a domain name is BBC. It will mark it as highly credible. On the other hand, if the website is not on the credible list, it will mark it as unknown. The second parameter of the programmed-based approach is checking the grammar. It checks the percentage of grammatical errors in the text using'gingerit' library. Lastly, the system will also check the count of swear words in the article. On the basis of these five results, which include two AI-based approaches and a program-based approach which is using three parameters, a final result will be taken in the form of a percentage. This whole procedure can be depicted in Fig. 1.

The rest of the draft is organized as follows; Section 2 contains reviewed literature; data and methodology is positioned in Section 3. The results of their discussion are part of Section 4, while the conclusion is ordered in Section 5.

2 Literature review

An extensive survey of contemporary scholarly work unveiled that the issue of fake news has been a major issue amongst scholars from various backgrounds. For instance, some researchers have witnessed that fake news is no longer a domain of the marketing and public relations departments [20]. Instead, the problem is escalating, being observed as

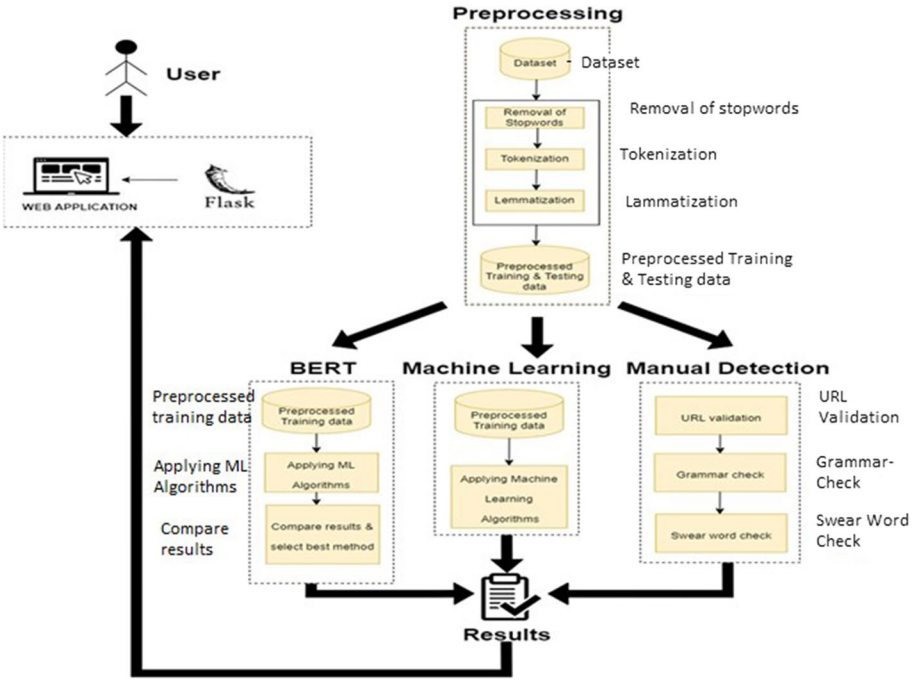

Fig. 1 System Flow

part of the responsibilities related to the information technology (IT) department. Conventionally, it was sup- posed that the two departments mentioned above were the ones to deal with any consequences arising from disseminating misleading news related to an organization. However, contemporary investigation specifies that fake news is measured to be a serious threat to information security. Therefore, the IT department's involvement is premised on the idea that it would help avoid the various risks linked with the problem.

This section will illustrate the researches that have already been done in detecting Fake news using different machine-learning techniques. Most of these researches focused on detecting fake news in online mediums like news articles, blogs, and tweets. These researches are not only limited to the English language but also cover the Urdu language. Researchers suggest that earlier Fake news systems work based on the binary classification and cannot tell how much fake news is directly related or unrelated to real news. To solve this, Dense Neural Network (DNN) is used with the feature extraction technique TF-IDF and the dataset used is "Emergent Dataset" by Craig Silverman [7]. TF-IDF techniques describe how the Naive Bayes Classifier, Support Vector Machine (SVM), and logistic regression algorithms are used on two different datasets (ISOT fake news dataset and Liar- Liar dataset). After data pre-processing, features are extracted using TF and TF-IDF. The research concludes that Naïve Bayes gives the best accuracy for the TF feature, and Logistic Regression gives the best accuracy, highest F1 score, precision, and recall for the TF-IDF feature. The research proposes that ISOT fake news dataset provides better results when SVM, Naïve Bayes, and Logistic regression models are performed on it for both TF and TF-IDF features [8].

Natural Language Processing approach can detect fake news using feature extraction TF-IDF along with the DrQA system and Linguistic Inquiry and Word Count (LIWC).

Three datasets have been used and are collected from various sources like Politifact, Channel4, Snopes, and social media. SVM, logistic regression, RNN, CNN, and LSTM models were applied. The research concludes the results and accuracy of these machine learning models [9].

In the detection of fake news through an ensemble voting classifier system along with the utilization of algorithms like Naive Bayes, K-NN, SVM, Random Forest, Logistic Regression, Gradient Boosting, etc., the data set consisting of a combination of both Fake and True news is used. In preprocessing, NLP is used to improve accuracy. There are also future works from this approach.

Integration of K-means clustering and SVM detected fake news effectively. The data sets used were Buzz Feed News, DS Detector, and LIAR. Decision tree and Naïve Bayer's approach would classify data according to occurrence and probability—every classifier on each data set performed above 90% [10]. Fake propaganda in political events such as Brexit and U.S. Presidential Elections in 2016 was spread through online articles. Semantic features of machine learning were used to spot these issues. RNN has been compared to Random Forrest and Naïve Bayes classifier, and the dataset is from Kaggle. Random forest classifier, using the bigram feature, has achieved 95.66% accuracy [11].

A new classification model can differentiate between real and fake news with an 83 percent accuracy. In this research, a dataset of both fake and real news is taken from Kaggle. Punkt articulation tokenizer performs data pre-processing and extracting features. Logistic Regression, SVM, and Naïve Bayes are used with Lidstone smoothing models for classification and regression. Lidstone smoothing is used to ensure that no probability will be zero. Naïve Bayes Classifier with Lidstone smoothing model achieves maximum accuracy [12].

FakeNewsTracker [13] can automatically collect and detect fake news using a deep-learning approach. The real and fake news dataset is gathered from different fact-checking websites and tweets. The social Article Fusion model is used to detect fake news. Feature extraction is done by learning news content and social engagement of the user with the news. Autoencoders are used to learn news content, and RNN, and LSTM are used to find how the user can interact with real and fake news. Features are extracted from the SAF network, and SVM, Naive Bays Classifier, and Logistic Regression models are used to perform classification.

A system detects [14] fake news using classification algorithms and Twitter reviews with a dataset from Twitter Streaming API, and data preprocessing is done along with this. Then for the classification task, the Naïve Bayes classifier is a machine learning algorithm and purposes Vectorization algorithm and NLTK (Natural Language Toolkit) for sentiment analysis and Twitter review analysis.

Five Machine learning models [15] Decision Tree, Naive Bayes, Logistic Regression, SVM, and Neural Networks along with TF-IDF and Count Vectorizer work on a given dataset. In this research, the dataset is created from the fake news dataset by Kaggle and the dataset generated by Andrew Thompson. Feature extraction is performed using the Bag of Words, and this model is implemented using Count Vectorizer and TF-IDF vectorizer. For the classi- fication task, Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machines, and Neural Networks models are used in this research. The SVM model and the TF-IDF vectorizer give the highest accuracy on a given dataset. Logistic regression gives good results, the Naive Bayes classifier works well on a small dataset, and Neural Networks and Decision Trees do not work well on a given dataset.

Fake news detection during the Covid 19 pandemic is done through NLP, word embedding, and LSTM. The data set was taken from the Covid-19 dataset by Sumit [16]. One-hot

encoding is used for word embedding. Training is done on 80% of the data with validation, and testing is done on 20% of the data. The data set achieved 96% accuracy.

Machine Learning and NLP techniques for detecting fake news use key expressions of news affairs and are used as the dataset [17]. Data is preprocessed. Stance is detected using the stance detection model. TF-IDF model and how we can extract features using the TF-IDF model. The training dataset uses the Passive-Aggressive Classifier, Multinomial Naive Bayes classifier, and SVM. Results are analyzed using Confusion Matrix, Accuracy Score, Precision, and Recall. Passive Aggressive with SVM gives us good accuracy scores and precision. Multinomial Naive Bayes gives us the highest recall.

Detection of fake news using Naive Bayes, Passive Aggressive classifier, and SVM through NLP techniques is done [18]. The dataset is from Kaggle and Signal media news. Naive Bayes, Passive Aggressive classifier, SVM, and Logistic regression models are discussed. Data pre-processing is performed using tokenization, stemming techniques, and stop words. Features are extracted through NLTK, Count, and TF-IDF vectorizer. The research proposes a Passive-Aggressive classifier model for detecting fake news as it gives the highest accuracy, up to 93%.

The research discussed the automatic detection of misinformation [19]. Two different data sets are created. Features are extracted using the Bag of Words, Linguistic Inquiry Word Count Software (LIWC), and the Stanford Parser tool. SVM and five-fold validation are used for classification.

Research conducted in Urdu suggests that fake news detection revealed that individual machine learning algorithms, combined with the ensemble learning technique, give better precise output in detecting fake news than individual machine learning algorithms [20].

The k-nearest neighbor (KNN) model measures fake news by taking data from 150 people's WhatsApp, Twitter, Facebook, etc. Metrics like Cohen Kappa etc., are used for prediction evaluation. The models distinguish the false news having a negative effect and discriminate against false news with a positive effect.

Fake news detection by Mining Dual Emotion [21] gives a concept of dual emotion. Dual emotion is the social emotion and the publisher emotion. The relationship between the integrity of news and dual emotional signals is de- fined. A feature set named dual emotion is used to expose distinctive emotional signals in detecting fake news. The proposed features can be plugged into existing detectors of fake news.

Linguistic features of misinformation based on titles and their statements discuss that the linguistic features are to be explored for distinguishing between real and fake statements news. For statements, Gradient Boosting gives 49.03% accuracy. For headlines and statements, the Logistic Regression gives 50.16% accuracy. For headlines, Extra Trees shows an accuracy of 77.57%.

The vulnerabilities of NLP techniques for fake news [22] raise concerns related to the fake news detectors, which are based on the linguistics aspects.

The experiments on the FakeBox show results that proved the effectiveness of fake news tampering in the existing fake news detection model. Existing models are vulnerable to fact tampering attacks, and these have biases. The solution for the problem is that methods should be extracting the key information from the reported article along with casual relationships and then comparing it with the news fact knowledge graph.

Detection of fake news spread via social media accounts [23] proposes sorting clickbait problems by identifying web pages that can mislead readers. The system accesses a webpage's bounce rate and then labels it as a potential fake news source if it is high. A comparison of classifiers based on F-measure, Recall, Precision, and ROC is made, and the highest precision achieved is 99.4%.

3 Data and methodology

This section describes the complete flow of research, from data collection to detecting fake news. It describes the dataset we used for training the model and preprocessing that data. It also describes feature extraction techniques, different Machine Learning models for training datasets, and manually programmed-based techniques we used to detect fake news.

3.1 Data collection

The system uses four different datasets according to four domains: English articles, English tweets, Urdu articles, and Urdu tweets. The datasets have two columns:' text' and'label'. The'label' column contains either'1' or'0' against the textual data of the'text' column. In these datasets,'1' is for the news, labeled as fake news, while'0' is for the news, labeled as real news. Table 1 explains the data instances inside the dataset. For English articles, we have total instances of about 20,800, out of which 10,413 are Fake, 10,387 are Real, and 39 are null values. For Urdu articles, we have total instances of about 900, out of which 400 are Fake, 499 are Real, and 1 has a null value. For Arabic articles, we have total instances of 4700, out of which 2230 are fake and 2470 are real, and 29 null values. For English tweets, we have cases total of about 8953, out of which 4693 are Fake, 4260 are Real, and 39 are Null values. The total number of tweets for Arabic Tweets was 7935, out of which 3750 were fake, 4185 were confirmed, and null values were 2. For Urdu articles, we have total instances of about 8953, of which 4693 are Fake, 4260 are Real, and 0 null values.

3.2 Data preprocessing

After the collection of data, there is cleaning and pre-processing of data. First, tokenization and removal of stop words are performed. After that, Lemmatization is implemented, rather than stemming to get the word to the root words.

3.2.1 Tokenization

For tokenization, the Punkt sentence tokenizer is implemented to split the plaintext of English, Arabic, and Urdu news articles into a list of sentences.

3.2.2 Stop words removal

As stop words do not affect the meaning of sentences, we remove stop words from both English, Arabic, and Urdu news datasets because they do not provide any important

Table 1 Dataset Information

Dataset	Instances	Fake news	Real news	Null values
English Articles	20,800	10,413	10,387	39
Urdu Articles	900	400	499	1
Arabic Articles	4700	2230	2470	29
English Tweets	8953	4693	4260	0
Urdu Tweets	8953	4693	4260	0
Arabic Tweets	7935	3750	4185	2

information about language to our model (Fig. 2). A sample of a few stop words is given in the Table 2.

3.2.3 Lemmatization

Wordnet Lemmatizer is used to convert a word into its meaningful root form by doing a morphological analysis of each word. After pre-processing, data is split into a 20–80 ratio. 20% data is testing data and 80% data is training data.

3.3 Feature extraction

For feature extraction, TF-IDF vectorization is used. The Term Frequency Inverse Document Frequency (TF-IDF) Vectorizer calculates the importance of words in a document and corpus. In TF-IDF vectorization calculation, term frequency values, and inverse

Stop Words	Language
And	English
We	English
The	English
اور	Urdu
ابہی	Urdu
آج	Urdu
و	Arabic
ال	Arabic
نحن	Arabic

Fig. 2 Sample of few stop words

Table 2 Result from classification algorithms for English news dataset

Algorithms	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.96	0.97	0.97	0.97
Multinomial Naïve Bayes	0.99	0.66	0.79	0.83
Gradient Boosting	0.92	0.95	0.94	0.93
Decision Tree	0.90	0.91	0.90	0.90
Random Forest	0.94	0.89	0.92	0.92
KNN	0.57	0.99	0.73	0.62

document frequency are calculated. Then the frequency value and inverse document frequency are aggregated using multiplication and normalization.

$$TF - IDF(t, d, D) = tf(t, d) * IDF(t, D) \tag{1}$$

Term Frequency (TF) counts the number of times a term appears in a document.
TF(t) = Number of times term $'t'$ appears in *a* document / Total number of terms in the document $'d'$
Document Frequency (DF) calculates the number of documents containing the term 't'.
IDF(t) = $\log_2$(Total number of documents (D)) / (Number of documents having term (t))

3.4 Classification

For the classification task, multiple techniques were applied for the detection of fake news. Machine learning algorithms, BERT, and various manually programmed techniques were used in the Classification step.

3.4.1 Machine learning algorithms

In the Machine learning approach, the following algorithms were used to perform classification tasks:

- Logistic Regression
- Multinomial Naïve Bayes
- Gradient Boosting
- Decision Tree
- Random Forrest
- K Nearest Neighbor

3.4.2 BERT (Bidirectional Encoder Representations from Transformers)

To increase the scope of our project, we used a deep learning model, based on the transformers called BERT (Bidirectional Encoder Representations from Transformers). Unlike other directional models, that can read the text in a sequence (from left to right or from right to left), it reads in both directions at once. First of all, Pretraining of BERT is implemented by adopting the following NLP strategies:

- MLM (Masked Language Model) hides a word in a sentence and then the surrounding words predict the word that has been hidden based on the context of the masked word. In MLM, 15% of the words have been masked in each sentence.

- NSP (Next Sentence Prediction) takes pair of sentences as input and predicts whether the second sentence has a sequential connection with the first sentence or not. By these two discussed techniques, BERT can deeply un- derstand the meaning of language. Then the Fine-tuning of BERT is done by using the classification layer on the transformer output. Pre-processing of text in BERT is done by combinations of embeddings. Figure 3 shows how pre-processing is implemented in BERT by position, segment, and token embeddings.

3.4.3 Manual programmed based techniques

In programmed-based detection, the system checks whether the news is fake or not on three parameters.

- First, the domain name is checked. This will check whether a domain name is of any highly credible news agency. For example, if a domain name is BBC. It will mark it as highly credible. On the other hand, if the website is not on the credible list, it will mark it as unknown.
- The second parameter of the program-based approach is checking the grammar. It checks the percentage of grammatical errors in the text using the'gingerit' library.
- Lastly, the system checks the count of swear words in the article.

4 Results and discussions

For machine learning algorithms, a confusion matrix is created with the values of True Negatives (TN), False positives (FP), False negatives (FN), and True Positives (TP).

Evaluation using the machine learning approach includes values of precision, recall, f1 score, and accuracy given by applied machine learning models.

$$Precision = True\ Positives\ (TP)/(True\ Positives\ (TP) + False\ Positives\ (FP)$$

$$Recall = True\ Positives\ (TP)/(True\ Positives(TP) + False\ Negatives\ (FN))$$

$$F1\ score = 2 * ((precision * recall)/(precision + recall)).$$

$$Accuracy = True\ Positives\ (TP) + True\ Negatives\ (TN)/True\ Positives\ (TP) + True\ Negative\ (TN) + False\ Positives\ (FP) + False\ Negatives\ (FN).$$

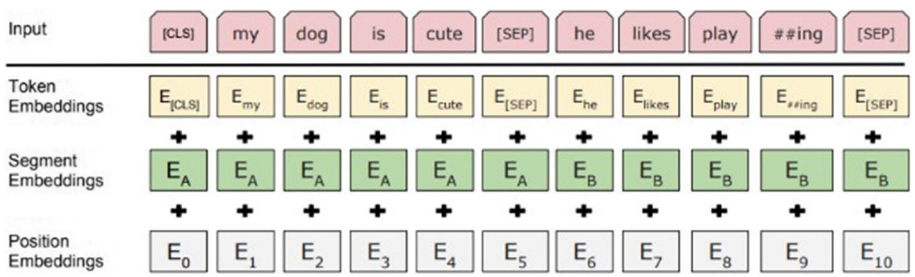

Fig. 3 Preprocessing in BERT (https://analyticsindiamag.com/a-guide-to-text-preprocessing-usingbert/)

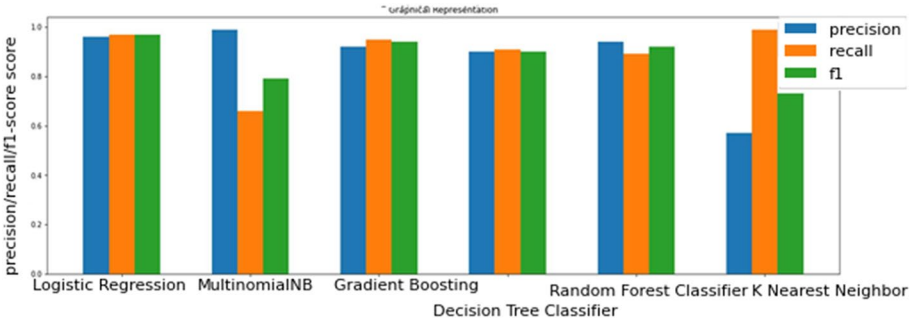

Fig. 4 Graph representing classification algorithms performance for English news dataset

Table 2 shows the performance of machine learning algorithms applied to the English news dataset.

After calculating these values, a bar chart is plotted for the values of precision, recall, and F-1 score at the y-axis with classification algorithms at the x-axis. The bar chart can be seen in Fig. 5. Logistic Regression has higher values in comparison to others. After Logistic Regression, Gradient Boosting, and Decision Tree can be seen for higher values of precision, recall, and F1- score. In the bar chart, there are three colors. Blue is representing precision, Orange is representing recall, and the F1 score is represented with green color (Fig. 4).

Now we evaluate the performance of machine learning algorithms on the English tweets dataset. The following Table 3 shows the performance of machine learning algorithms applied to English tweets dataset:

After calculating these values, a bar chart is plotted for the values of precision, recall, and F-1 score at the y-axis with classification algorithms at the x-axis. The bar chart can be seen in Fig. 5. In the bar chart, there are three colors. Blue is representing precision, Orange is representing recall, and the F1 score is represented with green color.

The evaluation of the performance of machine learning algorithms on the Urdu news dataset is conducted. Table 4 shows the performance of machine learning algorithms applied to the Urdu news dataset.

After calculating these values, a bar chart is plotted for the values of precision, recall, and F-1 score at the y-axis with classification algorithms at the x-axis. The bar chart can be seen in Fig. 6. In the bar chart, there are three colors. Blue is representing precision, Orange is representing recall, and the F1 score is represented with green color. Gradient Boosting gives us better results in comparison with other machine learning algorithms.

Table 3 Results from classification algorithms for English tweets dataset

Algorithms	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.91	0.88	0.90	0.89
Multinomial Naïve Bayes	0.86	0.93	0.89	0.89
Gradient Boosting	0.89	0.80	0.85	0.85
Decision Tree	0.84	0.81	0.83	0.82
Random Forest	0.93	0.85	0.88	0.89
KNN	0.88	0.86	0.87	0.86

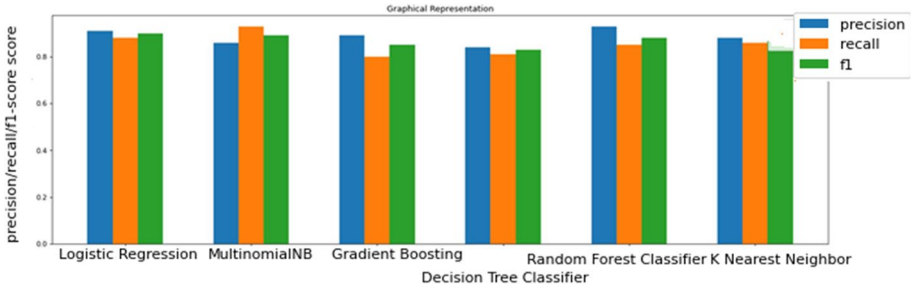

Fig. 5 Graph representing classification algorithms performance for English tweets dataset

Again we evaluate the performance of machine learning algorithms on the Urdu tweets dataset. The following Table 5 shows the performance of machine learning algorithms applied to the Urdu tweets dataset:

After calculating these values, a bar chart is plotted for the values of precision, recall, and F-1 score at the y-axis with classification algorithms at the x-axis. The bar chart can be seen in Fig. 7. In the bar chart, there are three colors. Blue is representing precision, Orange is representing recall, and the F1 score is represented with green color.

After training datasets on machine learning algorithms, we train English and Urdu news and tweets datasets on Transformer based algorithms i.e., BERT. The BERT model gives us accuracies of 90% and above on all 4 datasets. For final results and to detect whether the news or tweet is fake or real, the system follows different approaches in different scenarios. Different techniques according to their weightage, determine the credibility of news.

The following Table 6 shows the performance of machine learning algorithms applied to the Arabic news dataset.

After calculating these values, a bar chart is plotted for the values of precision, recall, and F-1 score at the y-axis with classification algorithms at the x-axis. The bar chart can be seen in Fig. 8. In the bar chart, there are three colors. Blue is representing precision, Orange is representing recall, and the F1 score is represented with green color. Gradient Boosting gives us better results in comparison with other machine learning algorithms.

The following Table 7 shows the performance of machine learning algorithms applied to the Arabic news dataset.

After calculating these values, a bar chart is plotted for the values of precision, recall, and F-1 score at the y-axis with classification algorithms at the x-axis. The bar chart can be seen in Fig. 9. In the bar chart, there are three colors. Blue is representing precision, Orange is representing recall, and the F1 score is represented with green color.

Figure 10 shows how the complete system works.

Table 4 Results from classification algorithms for Urdu news dataset

Algorithms	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.40	0.45	0.43	0.49
Multinomial Naïve Bayes	0.32	0.23	0.36	0.43
Gradient Boosting	0.55	0.55	0.55	0.62
Decision Tree	0.40	0.48	0.44	0.49
Random Forest	0.34	0.30	0.32	0.48
KNN	0.26	0.25	0.26	0.25

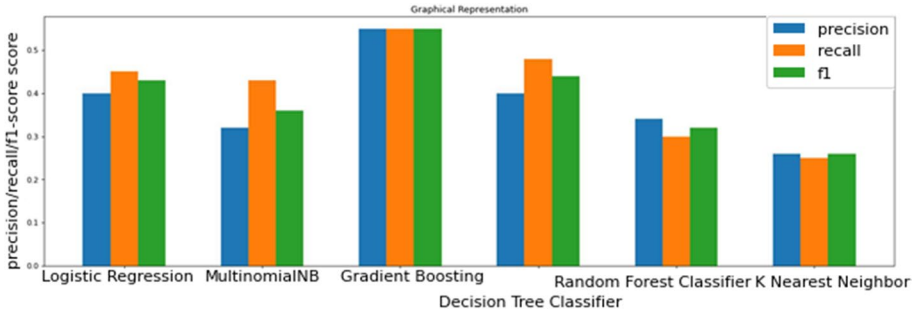

Fig. 6 Graph representing classification algorithms performance for Urdu news

The news article will have the URL of the source. So, for the prediction of news articles, weightage is given to each technique of prediction. If the URL of the news article is valid, the credibility of the news will be predicted according to the following weights:

- 15% of the weightage is given to the BERT algorithm e.g. If the news is fake according to the BERT algorithm 0 scores will be given out of 15 to input news.
- 15% of weightage is given to the best-performing machine learning algorithm in that domain of news e.g. If the news article is in the Urdu Language, the Gradient Boosting algorithm will check whether the news is fake or not and full score or 0 scores is given based on that result because Gradient Boosting algorithm performs best on Urdu news dataset.
- 45% is given to the URL of input news e.g. If the URL of input news is of valid source full score i.e., 45 out 45 will be given to that news otherwise 0 scores will be given.
- Grammar of news is weighted 5% e.g. If there are 5 or fewer grammar mistakes in the news article full score will be given to grammar of the news article and if there are 35 or more than 35 grammar mistakes in the news article 0 score will be given to grammar of news. So, the score of grammar is based on no. of grammatical mistakes in news articles.
- No. of swear words is weighted 10% e.g. If there are 10 or more swear words in a news article 0 score will be given out of 10 to that news article. Several swear words determine the 10 percent score of the news article.

If the URL of the news article is unknown, 30% of weightage is given to BERT. 30% of weightage is given to the best-performing machine learning algorithm on that domain of news.10% of weightage is given to URL and here news will achieve a 0 score out of 10 because of an unknown source. The grammar of news is weighted 15%. No. of swear

Table 5 Results from classification algorithms for Urdu tweets dataset

Algorithms	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.87	0.88	0.87	0.87
Multinomial Naïve Bayes	0.84	0.94	0.88	0.87
Gradient Boosting	0.89	0.80	0.84	0.84
Decision Tree	0.79	0.74	0.76	0.76
Random Forest	0.89	0.85	0.87	0.87
KNN	0.83	0.84	0.83	0.83

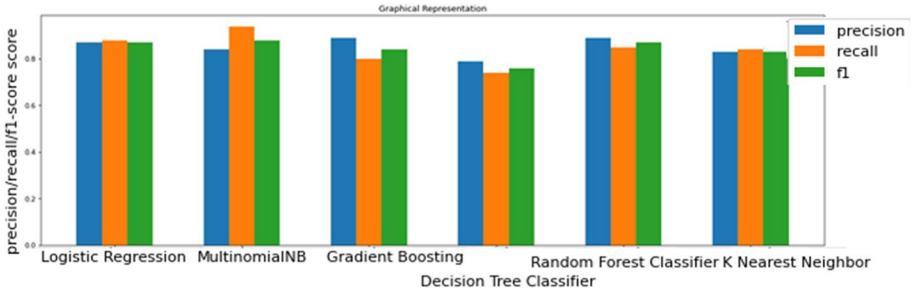

Fig. 7 Graph representing classification algorithms performance for Urdu tweets

Table 6 Results from classification algorithms for Arabic news dataset

Algorithms	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.30	0.35	0.33	0.39
Multinomial Naïve Bayes	0.43	0.43	0.26	0.33
Gradient Boosting	0.48	0.55	0.65	0.72
Decision Tree	0.54	0.58	0.64	0.54
Random Forest	0.44	0.40	0.42	0.38
KNN	0.26	0.35	0.36	0.35

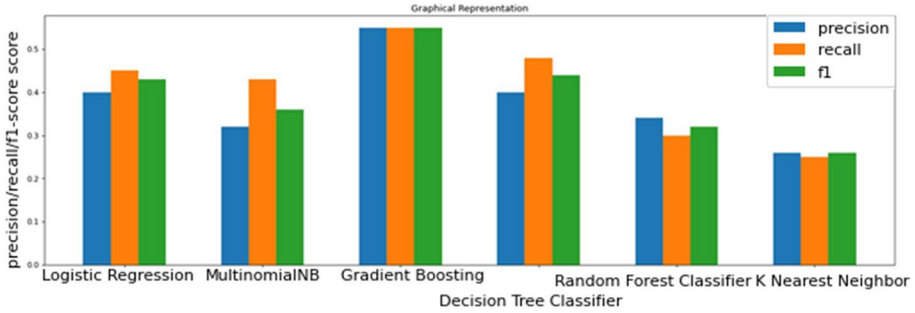

Fig. 8 Graph representing classification algorithms performance for Arabic news

Table 7 Results from classification algorithms for Arabic tweets dataset

Algorithms	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.77	0.78	0.83	0.82
Multinomial Naïve Bayes	0.87	0.89	0.84	0.81
Gradient Boosting	0.91	0.85	0.81	0.87
Decision Tree	0.89	0.77	0.71	0.79
Random Forest	0.79	0.78	0.79	0.80
KNN	0.73	0.71	0.74	0.78

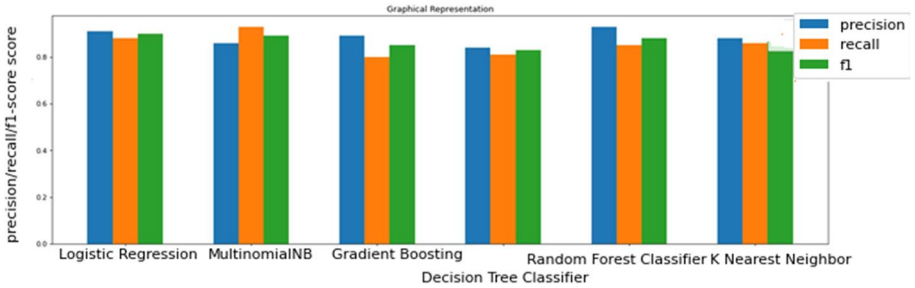

Fig. 9 Graph representing classification algorithms performance for English tweets dataset

words is weighted 15%. To check the credibility of the tweet, 30% of weightage is given to BERT. 30% of weightage is given to the best-performing machine learning algorithm on that domain of tweet (English or Urdu). The grammar of tweets is weighted 20%. The number of swear words in tweets is weighted 20%. The total score achieved by tweets out of 100 predicts the tweet is fake or real based on the combined result of different approaches. The total score achieved by a news or tweet out of 100 predicts the news or tweet is fake or real based on the combined result of different approaches.

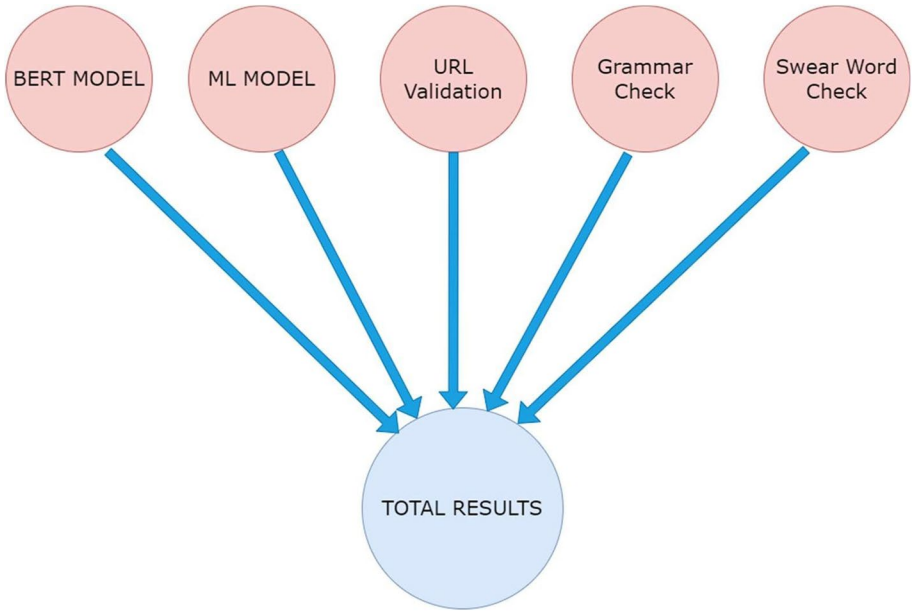

Fig. 10 The processes to find final results

5 Conclusion

This research focused to counter the spread of fake news. Hence the "Fake news detecting system". different machine learning techniques have been evaluated. The best-performing machine learning technique will be used in the web app along with the trained BERT model and manual-based program. Manual programs include different parameters such as URL validation, checking grammatical errors, and checking for swear words. The main purpose of this research is to propose a system that can detect the credibility of the news based on the textual content of the news. Different models of classification algorithms are implemented to identify the most suitable algorithm for the fake news detector system. The textual content can be used from online articles which are available in the form of blogs or news articles. Other than that, the textual data can also be used from tweets to check the credibility of that tweet. The system will be available in the form of a web app. The language for this textual content can be English, Arabic as well as Urdu.

In the future, the aim is to add many more functionalities. Initially, the main objective is to design a prototype and add tweet functionality. Upon completion of the final product for this research, we are motivated to deliver a fake news detection system dedicated to online news articles and tweets. That system would be able to detect news more efficiently and the spread of fake news can be reduced in the future. Other than that, in the future, we want to deliver a system that would be able to check the credibility of news from not just Twitter but also from other social media platforms such as Facebook, and Instagram.

Acknowledgements The authors extend their appreciation to the Deputyship for Research Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number:IFP22UQU4250002DSR230

Authors' contributions Atif Saeed, Muhammad Shahid Bhatti, Mohammed A. Al Ghamdi is the main author of the current paper. They contributed to the development of the ideas, design of the study, theory, result analysis, and paper writing. Zeeshan Gillani, Sultan H. Almotiri contributed to the result analysis and paper revision. All authors read and approved the final manuscript.

Data availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

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

References

1. Byers K (2021) How many blogs are there? (And 141 Other Blogging Stats). Growth Badger. [Online]. Available: https://growthbadger.com/blogstats. Accessed 30 12 2021
2. Silk J (2020) Spread of coronavirus fake news causes hundreds of deaths. DW. [Online]. Available: https://www.dw.com/en/coronavirusmisinformation/a-54529310. Accessed 30 12 2021
3. Bovet A, Makse HA (2019) Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 10(1):1–14
4. Menon AHS (2020) he dead professor and the vast pro-India disinformation campaign, BBC. [Online]. Available: https://www.bbc.com/news/world-asia-india-55232432. Accessed 30 12 2021
5. Fake News. Lies and Propaganda: How to Sort Fact from Fiction," M LIbraray, 2021 10 26. [Online]. Available: https://guides.lib.umich.edu/fakenews. Accessed 30 12 2021
6. B. UNIVERSITY, Fake News: Develop Your Fact-Checking Skills: What Kinds of Fake News Exist?," Benedictine University, 1 7 2021. [Online]. Available: https://researchguides.ben.edu/c.php?g=608230p=4219633. Accessed 30 12 2021

7. Thota A, Tilak P, Ahluwalia S, Lohia N (2018) Fake news detection: a deep learning approach. SMU Data Sci Rev 1(3):10
8. Acharya A, Jathan A, Anchan D, Dinesh Kottary K, Mr Sunil BN (2019) Fake news detection using machine learning
9. Ibrishimova MD, Li KF (2019) A machine learning approach to fake news detection using knowledge verification and natural language processing. In International Conference on Intelligent Networking and Collaborative Systems, pp. 223–234. Springer, Cham
10. Yazdi KM, Yazdi AM, Khodayi S, Hou J, Zhou W, Saedy S (2020) Improving fake news detection using k-means and support vector machine approaches. Int J Electron Commun Eng 14(2):38–42
11. Bharadwaj P, Shao Z (2019) Fake news detection with seman- tic features and text mining. Int J Nat Lang Comput (IJNLC) 8
12. Soni VD (2018) Prediction of geniunity of news using advanced machine learning and natural language processing algorithms. Int J Innov Res Sci Eng Technol 7(5):6349–6354
13. Shu K, Mahudeswaran D, Liu H (2019) FakeNewsTracker: a tool for fake news collection, detection, and visualization. Comput Math Organ Theory 25(1):60–71
14. Gurav S, Sase S, Shinde S, Wabale P, Hirve S (2019) Survey on automated system for fake news detection using NLP machine learning approach. Int Res J Eng Technol (IRJET) 6(01):308–309
15. Poddar K, Umadevi KS (2019) Comparison of various machine learning models for accurate detection of fake news. In 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), vol. 1, pp 1–5. IEEE
16. Islam MU, Hossain Md M, Kashem MA (2021) COVFake: A word embedding coupled with LSTM approach for COVID related fake news detection. Int J Comput Appl 975:8887
17. Khanam Z, Alwasel BN, Sirafi H, Rashid M (2021) Fake news detection using machine learning approaches. IOP Conf Ser Mater Sci Eng 1099(1):012040 (IOP Publishing)
18. Ahmed S, Hinkelmann K, Corradini F (2022) Development of fake news model using machine learning through natural language processing. arXiv preprint arXiv:2201.07489
19. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2017) Automatic detection of fake news. arXiv preprint arXiv:1708.07104
20. Swapna Y, Saiprasad P, Vathsalya B, Chandrakanth S (2019) Fake News Detection using Naïve Bayes Classifier
21. Zhang X, Cao J, Li X, Sheng Q, Zhong L, Shu K (2021) Mining dual emotion for fake news detection. Proc Web Con 2021:3465–3476
22. Khurana U, Bachelor Opleiding Kunstmatige Intelligentie (2017) The linguistic features of fake news headlines and statements. Diss. Master's thesis, University of Amsterdam
23. Aldwairi M, Alwahedi A (2018) Detecting fake news in social media networks. Procedia Computer Science 141:215–222

Authors and Affiliations

Mohammed A. Al Ghamdi[1] · Muhammad Shahid Bhatti[2] · Atif Saeed[2] · Zeeshan Gillani[2] · Sultan H. Almotiri[1]

Mohammed A. Al Ghamdi
maeghamdi@uqu.edu.sa

Atif Saeed
asaeed@cuilahore.edu.pk

Zeeshan Gillani
zeeshangillani@cuilahore.edu.pk

Sultan H. Almotiri
shmotiri@uqu.edu.sa

[1] Computer Science Department, Umm Al-Qura University, Makkah, Saudi Arabia

[2] Department of CS, COMSATS University Islamabad Lahore, Lahore 54600, Pakistan