Machine Learning Models for Fake News Detection: A Comparative Analysis

Yogesh S. Modhe, D. B. Kshirsagar

Department of |Computer Engineering, Sanjivani College of Engineering, Savitribai Phule Pune University, Kopargaon 423603, Maharashtra, India.

Corresponding Author – y.modhe@gmail.com

HOW TO CITE:
Yogesh S. Modhe, D. B. Kshirsagar (2026). Machine Learning Models for Fake News Detection: A Comparative Analysis.
International Journal of Special Education, 41(7s), 411-436.

COPYRIGHT STATEMENT:
Copyright: © 2026 Authors.
Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:
In the context of this paper, the focus has been on considering the use of machine learning models for fake news which is a growing concern in the growing digital age. The study focuses on four models of the model Random Forest, AdaBoost, Support Vector Machine, and Long Short-Term Memory to assess the accuracy of classification of the news articles as fake or actual. A fairly considerable volume of data was chosen to train and validate the models, and their efficiency was evaluated with metrics of classification accuracy. The findings show that the model we used for classification, Random Forest, achieved the highest level of accuracy, 76.43% when used to differentiate between fake and real articles from the other three models. Thus, the findings indicate that ML, especially Random Forest, can serve as a useful instrument used to counterfactual the phenomenon of fake news and misinformation by developing an automatic classification method for news articles. The following has been pointed out that these models can benefit media, social networks, and independent checkers as different sides of the information war against the issue. Therefore, the research advances the existing knowledge in the area of the development of automated systems for the classification of fake news.

Keywords: Fake news, Machine learning, Accuracy identification, Random Forest, AdaBoost model, SVM (Support Vector Machine), LSTM (Long Short-Term Memory), Classification problem, Dataset gathering, Misinformation

1. Introduction

Misinformation is one of the gravest problems of the modern world, which has been amplified by the growth of the new generation of social media and news websites all over the world. Despite these platforms' potential for openness and exchange of information, it opens a way for fake news, also known as "false news". Fake news can be defined as news that has been forged in such a way that relates untruths as believable news to make people trust the transmitted information to enable somebody to have his or her way politically, socially, or even financially [7]. The dissemination of such information is very dangerous to the integrity of real news sources, affects the media's credibility, and promotes social disruption. Thus, creating methods for identifying and preventing fake news became an important area of study this year, given the increasing focus on AI and ML-based solutions.

Therefore, deep learning which is a branch of machine learning has become more plausible in natural language processing (NLP) applications inclusive of fake news detection. Statistical information, articles, and news also contain text and multimedia content which when analyzed can be worked upon to train deep learning models capable of identifying patterns that sort between truth and falsehoods [49, 50]. However, the issues of accuracy with such systems can be a problem since they are designed to largely mimic legit articles in the form of fake news and tend to copy these from real ones.

The high speed at which fake news spreads affects politics provides wrong information to the public during crises and controls the thoughts of the public on politics across the world [19]. Current methods of fake news detection always face challenges of real-time detection and categorization because of the different writing styles and sources of the fake news. Further, the generation of multimedia-based content, inclusive of pictures and videos, plus infographics, has compounded the constraints and complications of accurately identifying fake news [16]. Popular fake news detection tools, which can be described as simple keyword-based matching and fact-checking, are useless when it comes to these new media formats.

However, the present endeavor consists of developing an improved fake news detection model capable of addressing the complexity and diversification of information flow in real-time mode. Multimedia input must be incorporated into the solution and simple classification methods must be replaced by superior methods [51, 52]. Furthermore, the system should take into consideration the rate with which fake news circulates in the community for purposes of early intervention to counter its wide circulation.

The study aims to Design & Develop an Improved Fake News Detection System using deep learning techniques. The main objectives includes to review and analyze the existing systems for fake news detection, to process Multimedia I/P into text using suitable techniques for extracting text relevant to fake news. Also the study aims to apply different classifiers to the extracted text relevant to fake news and compare its performance by designing and developing an efficient classifier as a hybrid model using suitable datasets and algorithms [53]. The models will be evaluated using metrics and compare it with existing systems.

The increase of fake news as a type of information affects decision-making, trust, and indeed the democratic processes. Current detection mechanisms are not efficient in processing and identifying the mids and variety of information present today with the introduction of image and video information. Modern methods include fact-checking and keyword matching in a manner that is resource-consuming and can potentially generate result inaccuracies [14, 54]. To this end, this study seeks to fill these gaps by building a real-time, deep learning-based system that can process text and multimedia at once. With enhanced detection rate and performance, the proposed system will provide a better solution in combating the problem of fake news dissemination to the common people as well as several organizations in an efficient manner.

2. Literature Review

Fake news or, generally, the detection of undesirable information has been an active research area in the last several years as it is relevant to political, social, and economic processes. The worry of fake news spreading has made researchers over time employ some techniques as well as algorithms that enhance identification. This literature review deals with the current strategies in fake news detection with an emphasis on NLP, deep learning practices, and multimedia analysis. The foundational methods of these techniques and the related mathematical formulation and models will also be explained.

2.1 Text-Based Fake News Detection Methods

Using textual data, many solutions to the fake news identification task have been proposed by different systems. The conventional strategies, in general, concentrate on detecting a selection of vocabulary characteristics like word frequency, sentiment, or syntactic patterns [12, 55]. In conventional solutions, such techniques as term frequency-inverse document frequency and bag-of-words have been used. Precisely, the mathematical formula of TD-IDF is the following TD-IDF: $\text{TD-IDF} = \text{TF} \times \log N/\text{DF}$, where TF represents the frequency of term T in document D, N is the total of documents, and DF is the number of documents containing the term T [56]. However, the conventional models have had limitations in the aspects of characterizing the semantic relationships between words. As a result, such advanced models as word embeddings [

4], such as Word2Vec and GloVe, began to be developed.

TF - IDF (t,d) = TF (t, d) x log ($\frac{N}{DF(t)}$)

$$P(w_{t-1}, w_{t+1}) = \frac{\sum_{i=1}^{V} \quad exp(v_{w_i} \ . v_c)}{exp(v_{w_t}. v_c)}$$

An example is given by Word2Vec, which uses a continuous bag-of-words or skip-gram model to maximize the likelihood of neighboring context words for a given target word. The equation of probability is represented as P(wt | wt−1, wt+1) = exp(vwt. vc) / $\sum V$ exp(vwi. vc), where vwt is the target word vector, vc is the context word vector, and V is the size of the vocabulary. Word embeddings help to map words onto a vector space, which is a great technique to capture their contextual relationships and improve the detection process. In that way, such methods as Word2Vec can provide these improved results as opposed to the older approaches by getting closer to the true meaning of words about fake news.

2.2 Deep Learning in Fake News Detection

Since LSTM networks and CNNs refer to advanced learning methods, they appear to be quite effective in identifying fake news. It is achieved by capturing important patterns in textual information, which is possible due to the features of LSTMs. Due to the special gating system, it is highly effective in preserving long-term connections and information flows [60]. In addition to mechanisms for forgetting, special information, and output gates are used to control the necessary data flow. The use of such processes guarantees the preservation of the most important request in the network for a long sequence of texts, proving that LSTMs are the most effective in identifying subtle patterns of liar detection.

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i)$$

$$\underline{C_t} = tanh(W_C . [h_{t-1}, x_t] + b_c)$$

$$C_t = f_t . C_{t-1} + i_t . \underline{C_t}$$

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o \ )$$

$$h_t = o_t . tanh(C_t)$$

It's common to utilize Convolutional Neural Networks in text classification tasks for the purpose of detecting fake news. CNNs make use of a sequence of convolutional steps to establish features from the given text. They prove to be relatively effective at going about identifying the patterns in n-grams since they can catch local dependencies of the text in question [59]. This drawing of complex data representations from data that is mixed finds a place in the classification problem domain. Therefore, both LSTM and CNN models are equally pivotal in the automatic detection of fake news.

2.3 Hybrid Models for Enhanced Accuracy

In recent years a piece of growing research that has caught the attention of many writers is the detection of misinformation and manipulated content, frequently referred to as fake news [61]. Naturally, as technology keeps progressing, these seemingly fake models also become more complex and believable. Recent approaches make use of hybrid models that combine several techniques to improve accuracy. Specifically, many of them take advantage of a combination of neural networks, each of them taking the best from what they do [7]. For instance, an LSTM network paired with a CNN could work for detecting fake news, as the former recognizes long-range connections in the data, such as sequences of text, and the latter extracts short-range patterns, such as n-gram features. Working alongside, the two networks can enhance the classification process by leveraging their respective advantages.

Loss Function:

$$L = -\sum_{i=1}^{n} \quad y_i \, log(\hat{y_i})$$

Prediction Formula:

$$\hat{y} = \sum_{i=1}^{n} \quad \alpha_i \hat{y_i}$$

For these blended models, categorization generally requires lessening an objective function, such as the cross-entropy loss, through the aid of which the model starts understanding the right relationships with input features and their labels. In addition, the use of ensemble learning approaches is highly frequent in the domain of fake news detection. Several classifiers are trained using the ensemble method, and their results are combined [57]. Several techniques exist to combine the results, and some of them include majority voting or weighted mean. The classifiers can have different levels of contributions depending on their weights [13, 58]. Through the ensemble methods, there is a benefit occurring by reducing the chance of mistakenly classifying the input. The resistance and overall accuracy are generally increasing as a result of integrating different models. By combining the results of multiple classifiers, the ability of the system to identify fake news is enhanced significantly in terms of both accuracy and reliability, particularly for large and complex datasets.

2.4 Multimedia Analysis for Fake News Detection

As multimedia content becomes more spread of fake news presented in images, videos, and other visual forms makes it hardly possible to identify in such a flow of multimedia data [11, 62]. It requires a comprehensive approach that is capable of analyzing this multimodal information and dealing with several types of content simultaneously. In this context, deep learning models are now indispensable tools, and Convolutional Neural Networks in particular are crucial for image analysis [63]. CNNs are designed to independently learn spatial hierarchies of features from input images, which helps them to recognize patterns and structures that suggest whether an image is associated with fake news or not.

$$f(I) = Pooling(ReLU(Conv(I)))$$

First thing, CNN receives an image and then uses convolutional layers to preserve important features. In general, every convolutional layer leverages filters to highlight definite patterns, such as edges or textures, while keeping the information about spatial connections between pixels. Activation functions, like ReLU, make feature extraction procedures better because of their ability to introduce non-linearity to the model, which allows better recognition of intricate relationships in the provided data. Then, we can usually use pooling operations to decrease the size of the feature maps. Dimensionality reduction is important not only to reduce the size of the calculation and make the model more effective but also to keep valuable data [6]. Having extracted and reduced valuable features, we can provide them to a classifier, which is usually a fully connected neural network. This classifier analyzes the extracted features to identify the probability of the provided image and misinformation being linked [63, 64]. The combination of these sophisticated procedures proves that CNN became sufficient at detecting deceiving and false information about multimedia materials. My point of view is that these procedures are going to become even more important for avoiding different types of misinformation as the way people share information is changing.

2.5 The Role of Social Network Analysis in Fake News Detection

Social network analysis (SNA) is critical in fake news detection and categorization given the structural interaction settings that people engage in on social media. Considering that fake news disseminates through structures dissimilar from categories of genuine news, studying the propagation networks is useful [65]. Such networks refer to the user as the nodes, and the shares, comments, likes, etc. as the edges. Through constructing these networks, the researchers can detect inconsistencies in users' behavior and patterns of diffusion related to fake news. Another method used in SNA includes analyzing different centralities measures such as degree, betweenness, and closeness centralities in the hope of identifying main influencers who could possibly be involved in circulating false news [66, 67]. For instance, degree centrality quantifies the number of outward or inward ties a user (node) has.

$$C_D(v) = \frac{d(v)}{n-1}$$

Where $C_D(v)$ is the degree centrality of node v, $d(v)$ is the degree of node v, and n is the total number of nodes in the network. Users with a high degree of centrality are often hubs in the information diffusion process and can amplify the spread of fake news.

Another approach includes examining the community detection algorithms which are useful in identifying sets of users who are likely to deliver similar content. The spread of fake news occurs mainly within closed groups of people, thus, the potential source and consequences can be easily established [68]. The Girvan Newman algorithm is typically used for this purpose and the modularity Q of a network is defined below.

$$Q = \frac{1}{2m}\sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

Where A_{ij} represents the adjacency matrix, k_i and k_j are the degrees of nodes i and j, m is a number of edges and $\delta(c_i, c_j)$ is 1 if nodes i and j belong to the same community [69, 70]. By incorporating these techniques, social network analysis becomes a powerful tool for identifying fake news and understanding its propagation mechanisms.

2.6 Ethical Considerations and Bias in Fake News Detection Algorithms

Several other ethical problems concerning fake news detection through ML algorithms are present apart from the bias. The issues have to be resolved to make sure that the outcomes are fair and accurate, as well as trustworthy [71]. Such problems arise from the processes through which data is collected as well as the naming and categorization of data sets and the bias that is inherent in algorithms. Algorithmic bias is among the plausible examples of ethical considerations associated with AI implementation. This implies that bias can occur at the data selection, labeling, and feature extraction stages of the data mining process. For example, if the given data set is dominated by 'fake news' sources or a specific region, it is possible for the algorithm to classify genuine news from other sources or regions as fake. This may result in specific populations pointing to Discrimination which in turn exposes them to exclusion.

Another equally critical issue is over-diagnosis as well as under-diagnosis because both are likely to result in negative consequences. Labeling real news as fake results in the rejection of genuine sources of information while recognizing fake news as real (false negative) allows the spread of misinformation. In other words, this means that in order not to cause harm, there must be a balance between precision and recall values [70, 71]. Privacy concerns are also considered within ethical problems duality. Some of the machine learning models use data that is gathered from social media analysis of large groups that may involve personal information. In case of no appropriate precautions, such data may negatively be used or its use may have negative consequences like surveillance or violation of users' right to privacy. Table 1 Common ethical issues and their potential consequences

Table 1 Common ethical issues and their potential consequences

Ethical Challenge	Description	Potential Consequences
Algorithmic Bias	Bias in training data leads to unfair or skewed outcomes	Discrimination, unequal treatment of groups
False Positives/Negatives	Misclassification of real or fake news	Loss of credibility, spread of misinformation
Privacy Concerns	Use of personal data in detection algorithms	Invasion of privacy, data misuse
Accountability	Lack of transparency in algorithm decision-making	Inability to challenge or rectify decisions

However, it is imperative to understand that designating these ethical issues entails extra processes such as algorithm open access, having diverse data, and conducting constant checks and monitoring of the performances of the model to enhance fairness.

3. Methodological description

3.1 Research Design and Framework

This study uses exploratory research to this assignment uses exploratory research to enhance the fake news detection system in deep learning methodologies [72]. Exploratory research helps the researcher focus on some issues that make it very difficult to distinguish between the two [29]. The design aids this in comparing the machine and deep learning classifiers in terms of their usability and feasibility in solving the problem of fake news [4, 30, 31]. Therefore, this method constitutes finding the latent structure and conceptual environments that will permit the research to establish the best way of modeling. The design also permits one to combine many approaches and methodologies that are critical when creating a new and better hybrid model for improving the classification results and increasing misinformation detection ability.

3.2 Data Collection and Preprocessing Techniques

For data collection for this particular study, data was collected from Kaggle using a secondary data collection technique since it had a varied variety of news articles labeled. It includes n cases, of which the text content is an important feature, and other features include labels representing true (0) or fake (1) news. The goal was to have an equal distribution of the two classes to allow the model to learn well from the data features [32, 33]. The most important phase is the preprocessing phase, where the raw text data is converted to a suitable format for analysis. First, preprocessing of the text is performed with regards to changing the text case to lower and eliminating the punctuation marks.

The mathematical representation of text can be expressed as a bag-of-words (BoW) model, where a vector represents each document. V_i:

$$V_i = [f_1, f_2, f_3, \ldots\ldots f_n]$$

Where f_j represents the frequency of term j in the document D_i.

Moreover, tokenization breaks down the document into tokens, and stop-word elimination also eliminates words that do not significantly affect the content [2, 3, 34]. At last, after data analysis, the dataset is again transformed to a new vector form by using the TF-IDF technique to get the weighted vectors that help in extracting more features from the data and enhance the performance of the model.

The resultant feature matrix X can be defined as:

$$X_{ij} = TF - IDF(t_j, d_i)$$

Where t_j is the j^{th} term and d_i is the i^{th}Document.

3.3 Model Selection and Justification of Choices

In the current study, the use of a range of machine learning and deep learning models were utilized to create a sophisticated fake news detection model. The chosen models are logistic regression, random forest, support vector machines, AdaBoost, multilayer perceptron, Long Short Term Memory, and a combined model that incorporates the above-mentioned methods for better classification [35, 36].

Logistic Regression (LR) is the most basic statistical model that predicts the probability of a binary event occurrence. It employs the logistic function, represented mathematically as:

$$P(Y=1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}}$$

Such a model is justified given the interpretability and when used for binary classification tasks such as fake news identification.

Random Forest (RF) is a tree-based learning algorithm that combines many decision trees to enhance accuracy and reduce overfitting [4, 37]. The combined prediction is formulated as:

$$\hat{Y} = \frac{1}{T}\sum_{t=1}^{T} h_t(X)$$

Where T is the number of trees, enhancing predictive performance.

Support Vector Machine (SVM) deals mainly with selecting one of the hyperplanes with a maximum margin for separation, especially for high-dimensional datasets. Its decision function is expressed as:

$$f(X) = w^T X + b$$

Another ensemble concept is AdaBoost which builds weak classifiers sequentially, and at each step re-weights misclassified instances [38,39]. The final model is given by:

$$F(X) = \sum_{m=1}^{M} \alpha_m h_m(X)$$

Where $\alpha_m$represents the weight of each classifier.

Multi-Layer Perceptron (MLP) is a neural network model capable of learning complex representations, expressed as:

$$Y = f(W^T X + b)$$

Where f is an activation function.

Certainly, Long short-term memory (LSTM) networks are effective at modeling sequential dependencies inherent in textual data which is crucial for capturing context in fake news [6]. Its hidden state is computed as:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

Finally, the Hybrid Model combines outputs from multiple classifiers to leverage their strengths, represented by:

$$\hat{Y}_{hybrid} = MetaModel(f_1(X), f_2(X), \ldots, f_n(X))$$

This systematic choice and inclusion of models has a very important contribution towards increasing the dependability and accuracy of fake news detection systems.

3.4 Data Analysis and Evaluation Methodology

During the modeling phase of the fake news detection process in analyzing data, some aspects had to be passed through to create the classification and the general assessment of the model. Firstly, the dataset was split into training and testing where 80% of the data was assigned as fake news data and the remaining 20% as real news data [5, 40]. This division also has the responsibility of checking that the model is well trained when this is being carried out while checking on the competency of the model on new data. Thus, to compare the results of the work the inclusion accuracy, precision, recall, and F1-measure were applied. Table 2 outlines the evaluation metrics used, offering insights into the criteria and methods applied for assessing performance or outcomes.

Table 2 Evaluation metrics

Metric	Description
Accuracy	The overall correctness of the model's predictions
Precision	The correctness of positive predictions
Recall	The model's ability to identify actual positives
F1-Score	A balance between precision and recall

These were complemented by confusion matrices to help in the determination of the classification outcome and cases of misclassification. This made it possible to provide a comprehensive assessment of the hybrid model thereby improving its ability to correctly classify the fake news and increasing its reliability.

Flowchart of the whole process

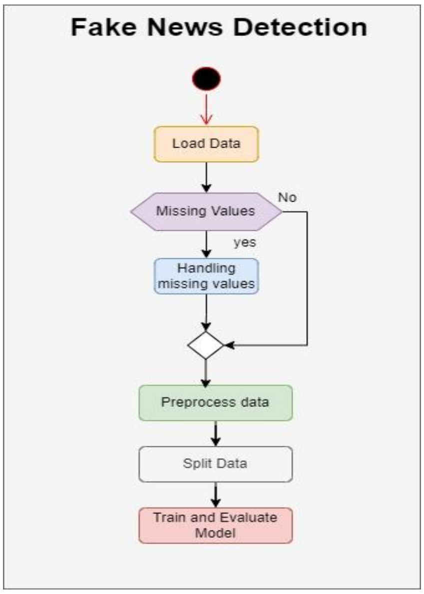

Figure 1 Flowchart of the analysis

Figure 1 outlines the key steps of the data analysis process, starting with loading and preprocessing the dataset. It includes handling missing values, splitting the data into training and testing sets, training various machine learning models, evaluating their performance, comparing results, and visualizing the outcomes for analysis.

4. Experimental Result

4.1 Data Loading and Exploration

Figure 2 Data loading

The code snippet includes necessary libraries and packages for data manipulation, data visualization, and machine learning algorithms. It uses a CSV file that involves news articles into a data frame and then, it prints the first five entries to help get a feel of the structure and data in the set. Figure 2 illustrates the process of data loading, showcasing the steps and mechanisms involved in preparing data for analysis or processing.

Get basic information about the dataset

```
3]: print(df.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2096 entries, 0 to 2095
Data columns (total 12 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   author                   2096 non-null   object
 1   published                2096 non-null   object
 2   title                    2096 non-null   object
 3   text                     2050 non-null   object
 4   language                 2095 non-null   object
 5   site_url                 2095 non-null   object
 6   main_img_url             2095 non-null   object
 7   type                     2095 non-null   object
 8   label                    2095 non-null   object
 9   title_without_stopwords  2094 non-null   object
 10  text_without_stopwords   2046 non-null   object
 11  hasImage                 2095 non-null   float64
dtypes: float64(1), object(11)
memory usage: 196.6+ KB
None
```

Figure 3 Basic information

Specifically, the `df.info()` method displays the DataFrame structure and reveals that the structure of the current DataFrame also has 2096 entries and 12 columns (Figure 3). It depicts data types and several occurrences of non-null values and memory usage in bytes showing that several original columns, concerning the contents of the articles and their characteristics, include missing entries.

Check for missing values

```
]: print(df.isnull().sum())

author                    0
published                 0
title                     0
text                     46
language                  1
site_url                  1
main_img_url              1
type                      1
label                     1
title_without_stopwords   2
text_without_stopwords   50
hasImage                  1
dtype: int64
```

Figure 4 Checking missing values

The `df. isnull(). sum()` function estimates the number of missing values in each of the columns of the data frame (Figure 4). It reveals that the 'text' feature has 46 entries missing completely, and the 'language', 'site_url', and 'text_without_stopwords' features also contain some missing values that describe incomplete data.

4.2 Data preprocessing

```
[5]: # Fill missing values for 'language' with the most common language
df['language'].fillna(df['language'].mode()[0], inplace=True)
# Fill missing values for 'site_url' and 'main_img_url' with a placeholder
df['site_url'].fillna('Unknown', inplace=True)
df['main_img_url'].fillna('No Image', inplace=True)
# Fill missing values in 'type' and 'label' with the most common type/label
df['type'].fillna(df['type'].mode()[0], inplace=True)
df['label'].fillna(df['label'].mode()[0], inplace=True)
# Fill missing 'title_without_stopwords' and 'text_without_stopwords' with the respective column
df['title_without_stopwords'].fillna(df['title'], inplace=True)
df['text_without_stopwords'].fillna(df['text'], inplace=True)
```

```
[6]: # Fill missing text data with an empty string
df['text'].fillna('', inplace=True)
df['text_without_stopwords'].fillna('', inplace=True)
```

Figure 5 Removing null values

The handling of missing values in the data frame is done by using several imputation techniques. The field 'language' is populated with the most commonly used language, and the 'site_url' and 'main_img_url' are replaced with dummy values. Furthermore, the 'type', 'label', 'title_without_stopwords', and 'text_without_stopwords' columns may contain missing values and these are appropriately replaced by suitable values or default values depending on the data analysis being conducted (Figure 5).

```
8]: # Check the distribution of the target variable
    print(df['label'].value_counts())

    label
    Fake    1295
    Real     801
    Name: count, dtype: int64

9]: # Encode the 'label' column as binary (e.g., 0 for 'Fake', 1 for 'Real')
    df['label'] = df['label'].apply(lambda x: 1 if x == 'Real' else 0)

0]: # Drop columns not needed for model training (e.g., URLs, image URLs)
    df.drop(columns=['main_img_url', 'site_url'], inplace=True)
```

Figure 6 Data Conversion

Figure 6 examines the distribution of the target variable, where we find 1,295 fake and 801 real news articles. The 'label' column is then converted to binary, where Real is represented by '1' and Fake by '0'. Irrelevant features, including URL links and image links, were stripped off to help construct the original training set for the model.

Data Split

```
]: # Split the data into features and target
   X = df['text_without_stopwords']
   y = df['label']

]: # Transform text data into TF-IDF features
   tfidf = TfidfVectorizer(max_features=5000)
   X_tfidf = tfidf.fit_transform(X)

]: # Split into train and test sets
   X_train, X_test, y_train, y_test = train_test_split(X_tfidf, y, test_size=0.2, random_state=42)
```

Figure 7 Data Preparation

The dataset is divided into features and target variables, with `X` representing the text and `y` the labels (Figure 7). The text data is transformed into TF-IDF features, limited to 5,000 maximum features. Finally, the data is split into training and testing sets, allocating 20% for testing.

4.3 Data Visualizations

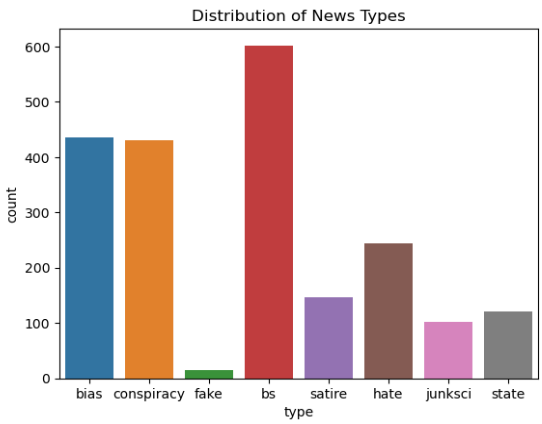

Figure 8 Distribution of news types

Figure 8 illustrates the distribution of news types within the dataset, revealing that "bs" (likely representing "bullshit" or sensational content) has the highest frequency. This suggests a significant prevalence of potentially misleading or exaggerated news articles compared to other categories in the dataset.

Figure 9 Word cloud

The word clouds visually represent the most frequent terms in fake and real news articles (Figure 9). In the fake news cloud, prominent words include "one," "Trump," "Hillary," and "Clinton," indicating a focus on controversial political figures. Conversely, the real news cloud highlights "people," "time," and "American," suggesting a broader, more general context.

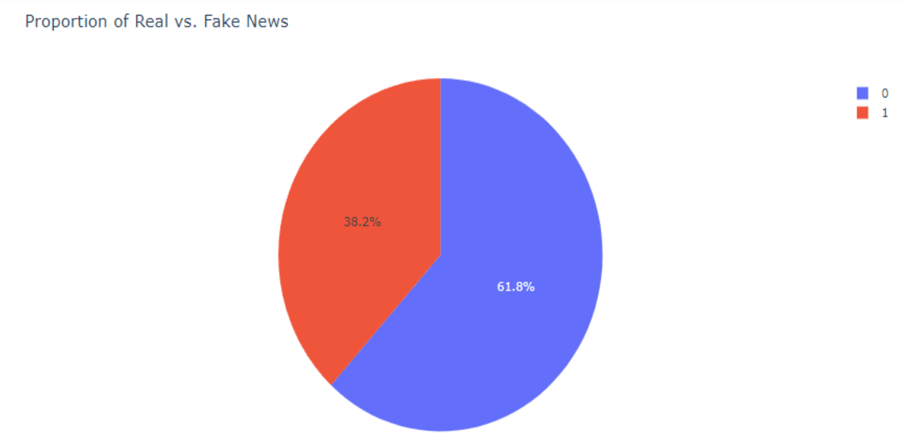

Figure 10 Proportion of Real vs. Fake News

Figure 10 illustrates the distribution of news articles labeled as real or fake. It reveals that 61.8% of the articles are categorized as fake news, while 38.2% are identified as real news. This significant prevalence of fake news highlights the need for effective detection and analysis methods.

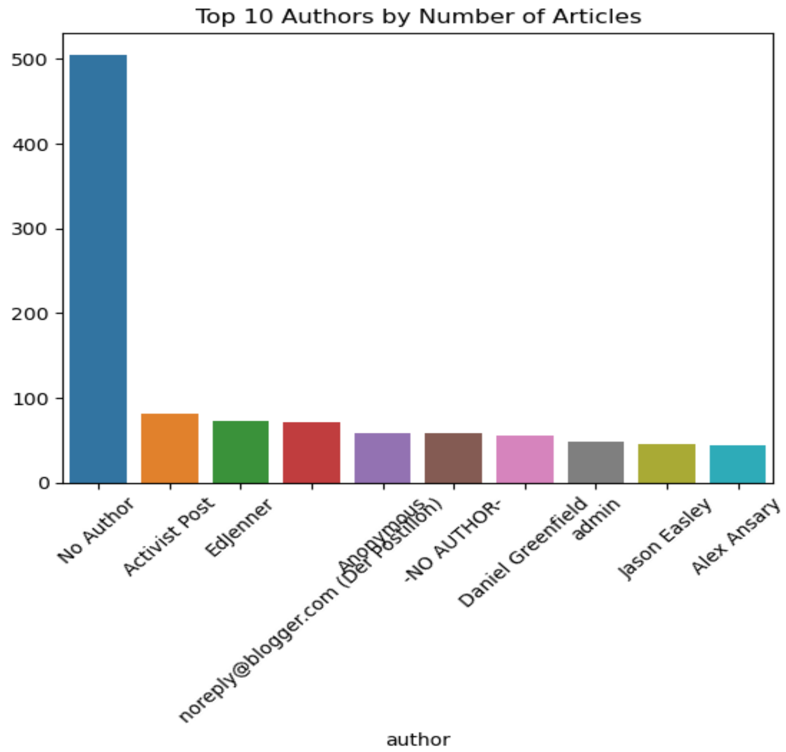

Figure 11 Top 10 Authors by Number of Articles

Figure 11 displays the top 10 authors based on the number of articles published. "No Author" has the highest count, indicating a significant amount of anonymous or uncredited content. Following that, "Activist Post" ranks second, suggesting a notable contribution from this specific source to the dataset.

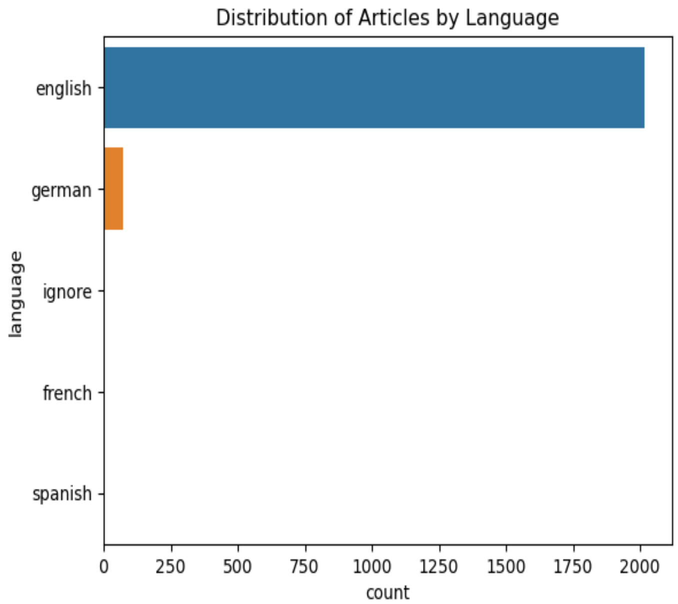

Figure 12 Articles by language

Figure 12 illustrates the distribution of articles by language, revealing that English dominates the dataset. This highlights a significant prevalence of English-language content, suggesting that the dataset primarily reflects news articles produced in English, which may influence the classification model's training and performance.

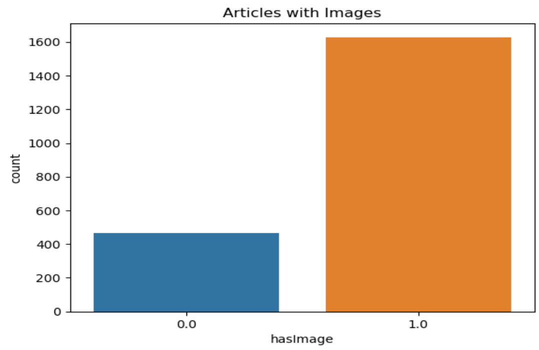

Figure 13 Articles with Images

Figure 13 plot displays the distribution of articles with images, indicating that the majority of articles contain images, labeled as yes. This suggests a strong trend in the dataset towards visual content, potentially enhancing the engagement and appeal of the articles to readers.

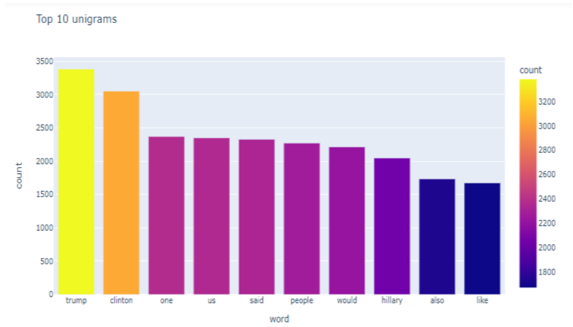

Figure 14 Top 10 unigrams

Figure 14 illustrates the top 10 unigrams in the dataset, highlighting that "Trump" has the highest frequency, followed closely by "Clinton." This indicates a strong association of these names with the news articles, potentially reflecting their prominence in political discussions and media coverage during the analyzed period.

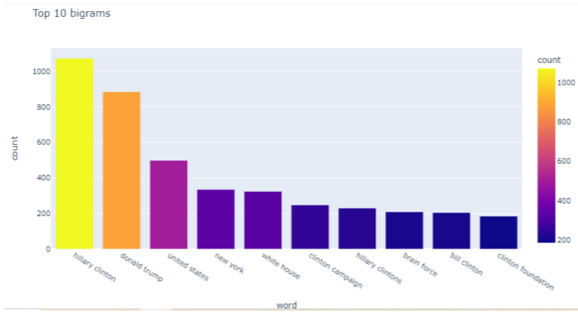

Figure 15 Top 10 bigrams

Figure 15 displays the top 10 bigrams from the dataset, prominently featuring "Hillary Clinton," "Donald Trump," and "United States." These phrases reflect significant political themes and key figures in the news articles, indicating the focus on U.S. politics and highlighting the relevance of these personalities in media discussions.

4.4 Model Creation and Evaluation

```
print("Logistic Regression Confusion Matrix:")
print(cm)
```

```
Logistic Regression Accuracy: 0.7404761904761905
              precision    recall  f1-score   support

           0       0.73      0.94      0.82       263
           1       0.80      0.41      0.54       157

    accuracy                           0.74       420
   macro avg       0.76      0.67      0.68       420
weighted avg       0.75      0.74      0.71       420

Logistic Regression Confusion Matrix:
[[247  16]
 [ 93  64]]
```

Figure 16 Logistic regression

The Logistic Regression model achieved an accuracy of approximately 74.05%. The classification report reveals a precision of 73% for fake news and 80% for real news, with a recall of 94% and 41%, respectively (Figure 16). The confusion matrix indicates 247 true negatives and 64 true positives, highlighting potential misclassification issues with real news.

```
y_pred_rf = rf.predict(X_test)
print("Random Forest Accuracy:", accuracy_score(y_test, y_pred_rf))
print(classification_report(y_test, y_pred_rf))
# Calculate the confusion matrix
cm = confusion_matrix(y_test, y_pred_rf)
print("Random Forest Confusion Matrix:")
print(cm)
```

```
Random Forest Accuracy: 0.7642857142857142
              precision    recall  f1-score   support

           0       0.77      0.88      0.82       263
           1       0.74      0.57      0.64       157

    accuracy                           0.76       420
   macro avg       0.76      0.72      0.73       420
weighted avg       0.76      0.76      0.76       420

Random Forest Confusion Matrix:
[[232  31]
 [ 68  89]]
```

Figure 17 Random forest

The Random Forest model achieved an accuracy of approximately 76.43%. The classification report shows a precision of 77% for fake news and 74% for real news, with recalls of 88% and 57%, respectively (Figure 17). The confusion matrix reveals 232 true negatives and 89 true positives, indicating balanced performance with some misclassifications.

```
svc.fit(X_train, y_train)
y_pred_svc = svc.predict(X_test)
print("SVM Accuracy:", accuracy_score(y_test, y_pred_svc))
print(classification_report(y_test, y_pred_svc))
# Calculate the confusion matrix
cm = confusion_matrix(y_test, y_pred_svc)
print("SVM Confusion Matrix:")
print(cm)
```

```
SVM Accuracy: 0.7452380952380953
              precision    recall  f1-score   support

           0       0.73      0.95      0.82       263
           1       0.82      0.41      0.54       157

    accuracy                           0.75       420
   macro avg       0.77      0.68      0.68       420
weighted avg       0.76      0.75      0.72       420

SVM Confusion Matrix:
[[249  14]
 [ 93  64]]
```

Figure 18 Support vector machine

The SVM model achieved an accuracy of approximately 74.52% (Figure 18). It demonstrated a precision of 73% for fake news and 82% for real news, with recalls of 95% and 41%, respectively. The confusion matrix indicates 249 true negatives and 64 true positives, highlighting strengths in identifying fake news but weaknesses in detecting real news.

```
ada = AdaBoostClassifier(n_estimators=50, random_state=42)
# Train the AdaBoost model
ada.fit(X_train, y_train)
# Make predictions
y_pred_ada = ada.predict(X_test)
# Evaluate the model
print("AdaBoost Accuracy:", accuracy_score(y_test, y_pred_ada))
print(classification_report(y_test, y_pred_ada))
# Calculate the confusion matrix
cm = confusion_matrix(y_test, y_pred_ada)
print("AdaBoost Confusion Matrix:")
print(cm)
```

```
AdaBoost Accuracy: 0.7571428571428571
              precision    recall  f1-score   support

           0       0.78      0.85      0.81       263
           1       0.71      0.60      0.65       157

    accuracy                           0.76       420
   macro avg       0.74      0.73      0.73       420
weighted avg       0.75      0.76      0.75       420

AdaBoost Confusion Matrix:
[[224  39]
 [ 63  94]]
```

Figure 19 Ada Boost Model

The AdaBoost model achieved an accuracy of approximately 75.71%. It recorded a precision of 78% for fake news and 71% for real news, with recalls of 85% and 60%, respectively (Figure 19). The confusion matrix reveals 224 true negatives and 94 true positives, indicating a balanced performance in classifying both categories.

```
# Train a Deep Learning model (Multi-layer Perceptron)
mlp = MLPClassifier(hidden_layer_sizes=(100,), max_iter=500, random_state=42)
mlp.fit(X_train, y_train)
y_pred_mlp = mlp.predict(X_test)
print("MLP Accuracy:", accuracy_score(y_test, y_pred_mlp))
print("\nClassification Report for MLP (Deep Learning):\n", classification_repor
cm = confusion_matrix(y_test, y_pred_mlp)
print("Multi-layer Perceptron Confusion Matrix:")
print(cm)
```

```
MLP Accuracy: 0.7452380952380953

Classification Report for MLP (Deep Learning):
              precision    recall  f1-score   support

           0       0.80      0.79      0.80       263
           1       0.66      0.67      0.66       157

    accuracy                           0.75       420
   macro avg       0.73      0.73      0.73       420
weighted avg       0.75      0.75      0.75       420

Multi-layer Perceptron Confusion Matrix:
[[208  55]
 [ 52 105]]
```

Figure 20 MLP Classifier

The Multi-layer Perceptron (MLP) achieved an accuracy of approximately 74.52%. It exhibited a precision of 80% for fake news and 66% for real news, with recalls of 79% and 67%, respectively (Figure 20). The confusion matrix indicates 208 true negatives and 105 true positives, demonstrating reasonable performance in distinguishing news types.

```
Epoch 2/5
21/21 [==============================] - 36s 2s/step - loss: 0.5203 - accuracy: 0.7134 -
4
Epoch 3/5
21/21 [==============================] - 35s 2s/step - loss: 0.2627 - accuracy: 0.9134 -
6
Epoch 4/5
21/21 [==============================] - 36s 2s/step - loss: 0.0931 - accuracy: 0.9679 -
6
Epoch 5/5
21/21 [==============================] - 38s 2s/step - loss: 0.0500 - accuracy: 0.9843 -
6
14/14 [==============================] - 4s 203ms/step

LSTM Model Accuracy: 0.7023809523809523

Classification Report for LSTM Model:
              precision    recall  f1-score   support

           0       0.77      0.76      0.76       263
           1       0.60      0.61      0.61       157

    accuracy                           0.70       420
   macro avg       0.68      0.68      0.68       420
weighted avg       0.70      0.70      0.70       420
```

Figure 21 LSTM Model

The LSTM model achieved an accuracy of approximately 70.24%. During training, accuracy improved significantly, peaking at 98.43% by the fifth epoch (Figure 21). However, the validation accuracy reached a maximum of only 69.94%. The classification report indicates a precision of 77% for fake news and 60% for real news, highlighting room for improvement, particularly in real news classification.

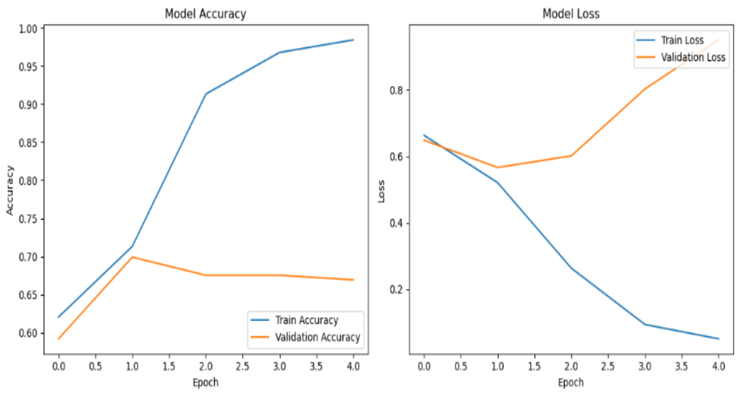

Figure 22 Model accuracy and loss

Figure 22 shows a significant increase in both training and validation accuracy over epochs, with training accuracy reaching approximately 98% while validation accuracy peaks around 70%. The loss plot demonstrates a decrease in both training and validation loss, although validation loss fluctuates, indicating potential overfitting as the model may not generalize well to unseen data.

```
21/21 [==============================] - 38s 2s/step - loss: 0.4130 - accuracy: 0.9067 - va
6
Epoch 4/5
21/21 [==============================] - 37s 2s/step - loss: 0.2123 - accuracy: 0.9463 - va
5
Epoch 5/5
21/21 [==============================] - 38s 2s/step - loss: 0.0810 - accuracy: 0.9791 - va
6
14/14 [==============================] - 5s 233ms/step
53/53 [==============================] - 11s 198ms/step
Stacking Hybrid Model Accuracy: 0.6714285714285714

Classification Report for Hybrid Model:
              precision    recall  f1-score   support

           0       0.70      0.83      0.76       263
           1       0.59      0.41      0.48       157

    accuracy                           0.67       420
   macro avg       0.64      0.62      0.62       420
weighted avg       0.66      0.67      0.66       420

Confusion Matrix for Hybrid Model:
[[218  45]
 [ 93  64]]
```

Figure 23 Stacking Hybrid Model

The stacking hybrid model achieved an accuracy of approximately 67%. The classification report indicates a precision of 70% for class 0 and 59% for class 1, reflecting better performance in identifying the negative class (Figure 23). The recall for class 1 is relatively low at 41%, suggesting that the model struggles to detect positive instances. The confusion matrix shows 218 true negatives and 64 true positives, with 93 false negatives, indicating room for improvement in predictions.

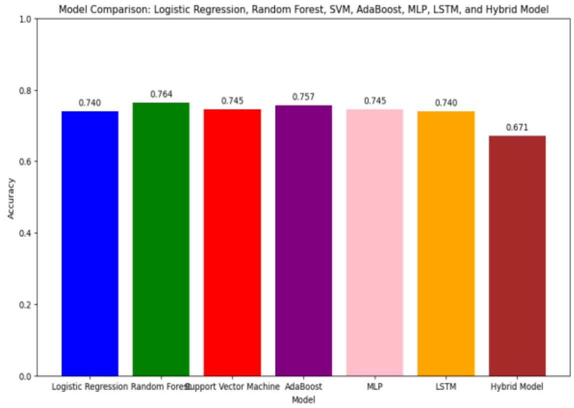

Figure 24 Model Comparison

Figure 24 compares the accuracies of various models, including Logistic Regression, Random Forest, Support Vector Machine, AdaBoost, MLP, LSTM, and the Hybrid Model. The highest accuracy is achieved by the Random Forest model, followed closely by AdaBoost. The LSTM and Hybrid models show moderate performance, while the MLP and SVM lag, indicating variability in model effectiveness for this task.

Table 3 Model result overview

Model	Model Type	Accuracy
Logistic Regression	Traditional ML	0.74
Random Forest	Traditional ML	0.764
Support Vector Machine	Traditional ML	0.745
AdaBoost	Ensemble Learning	0.757
Multi-layer Perceptron	Deep Learning	0.745
LSTM	Deep Learning	0.740
Hybrid Model	Ensemble Learning	0.671

Table 3 provides the details related to the accuracy scores of the categorization models, and their description for easy and enhanced understanding, and the models consist of Logistic Regression, Random Forest, SVM, AdaBoost, MLP, LSTM, and Hybrid Model.

5. Discussion

The performance evaluation of the different models during the analysis of the dataset of the news articles presents the important findings of how real and fake news can be detected with different methods highlighting the analysis strengths and weaknesses of the machine learning algorithms. The data loading and exploration section showed the richness of the data set by having 2096 records and 12 columns. This was because it was observed that certain columns had missing values and thus, it was deemed crucial to conduct proper data pre-processing for the analysis to yield accurate results [42, 43]. In pre-processing, some values were imputed for missing data and some columns were dropped to make the data set more suitable for models. The text data was preprocessed and converted into TF-IDF features that enabled correct representation in the subsequent models [7, 41]. As seen in the preceding visualizations, fake news constituted a significant portion, or 61.8% of the articles hence the need to set up efficient detection techniques. The model evaluation phase introduced the audience to several classifiers and it was clear that each of them had their specific advantages and limitations. However, the Random Forest model yielded higher results with an overall accuracy of about 76.43 %. Moderate precision and decent recall rates indicate high recognition of categories, although some unexpected classifications remained, especially with real news.

Logistic Regression and SVM both had accuracies of around 74%, SVM exhibited a high capability of fakeness identification but a poor ability to differentiate real news. The AdaBoost model also yielded a reasonable result of 75.71% which also supports that ensemble techniques classification can be improved. The hybrid model, which is observed to have the lowest performance with an accuracy of 67.10%, also displayed fairly good training performance with accuracy levels that have not been reflected in its validation. This implies that there might be an overfitting problem, which calls for precautions when tuning the model [44, 45]. All in all, this study supports the significance of utilizing sophisticated models and techniques to enhance the identification of fake news. Further research may consider using combinations of models of different types to increase their precision and minimize classification mistakes, which will further help to improve the classification of news in the context of the modern informational environment.

6. Limitations

Despite its notable contributions, the study has some limitations that may question the internal and external validity of the conclusions. First, the datasets mainly include the articles written in the English languages which brings a kind of language bias and restricted generalization to the other languages [45, 46]. Additionally, the case of missing values in certain features, especially target data, posed a challenge in this step, where data imputation could have skewed the final dataset. Furthermore, although different models of machine learning were discussed, there was no discussion on the possibility of the fine-tuning of what is called hyperparameters that would help to possibly enhance model performance. There was also a difference in the performance of the models and where the hybrid model seemed like it might overfit, as its training set had a higher accuracy than the validation set. This may imply that the model is not good at generalizing from one set of data to the next. One potential drawback is sample heterogeneity where not a wide range of sources are included in the dataset, and this can create a potentially narrow viewpoint in new articles.

7. Conclusion

In this paper, the development and assessment of multiple machine learning models for identifying fake news based on the data of news articles were made. The main considerations in conducting the research were the distribution of the different news types, the evaluation of the characteristics in the data, and the performance of the various models in differentiating the two categories real and fake news. Logistic Regression, Random Forest, SVM, AdaBoost, MLP, and LSTM were employed and their comparative analysis was done in terms of accuracy, precision, recall, and overall summary.

They proved that the Random Forest model was the most accurate model with an accuracy of 76.43% as compared to the AdaBoost model with an accuracy of 75.71%. In both models, the recall and precision for the model in classifying fake news and real news were reasonable. Both Logistic Regressions as well as the SVM models gave reasonable results with an accuracy of 74-75%, however, these models were slightly inaccurate while classifying real news articles as fake with more misclassifications. Both MLP and LSTM models had a high level of accuracy in the training data but the MLP had a surprisingly better result in handling the complex patterns in the unseen data while the LSTM model overfitted on the training data.

However, there are some challenges such as language bias, imbalance of the dataset, and many missing values were pointed out in the case of models used for the detection of fake news. It might be due to these factors that have affected the model's performance and its ability to generalize especially in the cases of real news detection, in which some of the models underperformed. It is important to note that the imputation strategies applied in handling missing data might also have brought in some error, and this might also be realized in the results. Altogether, the conducted analysis reveals the possibility of applying machine learning models to distinguish fake news with a relatively high degree of success. Nevertheless, widening the range of datasets used, optimizing hyperparameters, or applying more sophisticated methods can potentially boost model performance [47, 48]. Further research on fake news should be conducted to improve the approaches to fake news detection in different settings and different languages.

8. Acknowledgement

I would like to express my sincere gratitude to everyone who contributed to the successful completion of this research. Firstly, I am deeply thankful to my supervisor for their invaluable guidance, insightful feedback, and unwavering support throughout this study. I also wish to thank my family and friends for their continuous encouragement and understanding during this process. Lastly, I extend my appreciation to all participants and experts who contributed their time and knowledge, making this research possible.

9. Appendix

Model Overview

Model	Description	Accuracy	Precision (Fake News)	Recall (Fake News)	Precision (Real News)	Recall (Real News)
Logistic Regression	Simple, interpretable linear model used for binary classification.	74.05%	73%	94%	80%	41%
Random Forest	Ensemble model using multiple decision trees to improve performance and reduce overfitting.	76.43%	77%	88%	74%	57%
Support Vector Machine	Effective in high-dimensional spaces; uses a hyperplane to separate data points for classification.	74.52%	73%	95%	82%	41%
AdaBoost	Boosting method combining weak	75.71%	78%	85%	71%	60%

	learners to form a strong classifier.					
MLP Classifier	Multi-layer Perceptron, a type of neural network for classification tasks.	74.52%	80%	79%	66%	67%
LSTM Model	Recurrent neural network (RNN) is particularly useful for sequence prediction in text data.	70.24%	77%	77%	60%	60%
Hybrid Model	Stacking ensemble combining predictions from LSTM and traditional classifiers.	67.14%	70%	83%	59%	41%

References

Baarir, N.F. and Djeffal, A., 2021, February. Fake news detection using machine learning. In 2020 2nd International Workshop on human-centric smart Environments for health and well-being (IHSH) (pp. 125-130). IEEE. http://archives.univ-biskra.dz/bitstream/123456789/15806/1/Baarir_Nihel_Fatima.pdf

Shaikh, J. and Patil, R., 2020, December. Fake news detection using machine learning. In 2020 IEEE international symposium on sustainable energy, signal processing, and Cyber security (ESC) (pp. 1-5). IEEE. https://svu-naac.somaiya.edu/C3/DVV/3.4.5/Confernce+and+Book+Chapter/234.pdf

Ahmad, I., Yousaf, M., Yousaf, S. and Ahmad, M.O., 2020. Fake news detection using machine learning ensemble methods. Complexity, 2020(1), p.8885861. https://onlinelibrary.wiley.com/doi/pdf/10.1155/2020/8885861

Khanam, Z., Alwasel, B.N., Sirafi, H. and Rashid, M., 2021, March. Fake news detection using machine learning approaches. In IOP conference series: materials science and engineering (Vol. 1099, No. 1, p. 012040). IOP Publishing. https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012040/pdf

Sharma, U., Saran, S. and Patil, S.M., 2020. Fake news detection using machine learning algorithms. International Journal of Creative Research Thoughts (IJCRT), 8(6), pp.509-518. https://www.academia.edu/download/66254531/fake_news_detection_using_machine_IJERTCONV9IS03104.pdf

Ahmed, A.A.A., Aljabouh, A., Donepudi, P.K. and Choi, M.S., 2021. Detecting fake news using machine learning: A systematic literature review. arXiv preprint arXiv:2102.04458. https://arxiv.org/pdf/2102.04458

Mridha, M.F., Keya, A.J., Hamid, M.A., Monowar, M.M. and Rahman, M.S., 2021. A comprehensive review on fake news detection with deep learning. IEEE Access, 9, pp.156151-156170. https://ieeexplore.ieee.org/iel7/6287639/6514899/09620068.pdf

ASHFIA, J.K., MD ANWAR, H.W., MRIDHA, M.F., ALATIYYAH, M. and MD, A.H., 2022. AugFake-BERT: Handling Imbalance through Augmentation of Fake News Using BERT to Enhance the Performance of Fake News Classification. *Applied Sciences,* **12**(17), pp. 8398.

DEMENTIEVA, D., KUIMOV, M. and PANCHENKO, A., 2023. Multiverse: Multilingual Evidence for Fake News Detection. *Journal of Imaging,* **9**(4), pp. 77.

DIVYA, T.V. and MOHANTY, F., 2024. An Adaptive Word Vector Integrated With Enhanced Lstm For Classification Of Fake News. *Journal of Electrical Systems,* **20**(7), pp. 4062-4071.

HAUMAHU, J.P., PERMANA, S.D.H. and YADDARABULLAH, Y., 2021. Fake news classification for Indonesian news using Extreme Gradient Boosting (XGBoost). *IOP Conference Series.Materials Science and Engineering,* **1098**(5),.

K, S. and THILAGAM, P.S., 2023. Multi-layer perceptron-based fake news classification using knowledge base triples. *Applied Intelligence,* **53**(6), pp. 6276-6287.

LU, Y., SHEN, H., SHI, L., CHENG, N. and JIANG, H., 2023. An Explainable Fake News Analysis Method with Stance Information. *Electronics,* **12**(15), pp. 3367.

MAHAM, S., TARIQ, A., KHAN, M.U.G., ALAMRI, F.S., REHMAN, A. and SABA, T., 2024. ANN: adversarial news net for robust fake news classification. *Scientific Reports (Nature Publisher Group),* **14**(1), pp. 7897.

MORENO-VALLEJO, P., GISEL KATERINE BASTIDAS-GUACHO, MORENO-COSTALES, P. and JEFFERSON JOSE CHARIGUAMAN-CUJI, 2023. Fake News Classification Web Service for Spanish News by using Artificial Neural Networks. *International Journal of Advanced Computer Science and Applications,* **14**(3),.

NAGY, K. and KAPUSTA, J., 2021. Improving fake news classification using dependency grammar. *PLoS One,* **16**(9),.

NURENI, A.A., MISRA, S., DAVIDSON, O.O. and ABIDOYE, A.P., 2024. A Predictive Model for Benchmarking the Performance of Algorithms for Fake and Counterfeit News Classification in Global Networks. *Sensors,* **24**(17), pp. 5817.

PDF, 2024. MR-FNC: A Fake News Classification Model to Mitigate Racism. *International Journal of Advanced Computer Science and Applications,* **15**(2),.

RAHMANIAN, E., 2023. Fake news: a classification proposal and a future research agenda. *Spanish Journal of Marketing - ESIC,* **27**(1), pp. 60-78.

SANDU, A., IOANĂș, I., DELCEA, C., FLORESCU, M. and LIVIU-ADRIAN COTFAS, 2024. Numbers Do Not Lie: A Bibliometric Examination of Machine Learning Techniques in Fake News Research. *Algorithms,* **17**(2), pp. 70.

SHUSHKEVICH, E., ALEXANDROV, M. and CARDIFF, J., 2023. Improving Multiclass Classification of Fake News Using BERT-Based Models and ChatGPT-Augmented Data. *Inventions,* **8**(5), pp. 112.

A

Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, *4*, 100032.

Abdulrahman, A., & Baykara, M. (2020, December). Fake news detection using machine learning and deep learning algorithms. In *2020 international conference on advanced science and engineering (ICOASE)* (pp. 18-23). IEEE.

Mehta, D., Dwivedi, A., Patra, A., & Anand Kumar, M. (2021). A transformer-based architecture for fake news classification. *Social network analysis and mining*, *11*, 1-12.

Rohera, D., Shethna, H., Patel, K., Thakker, U., Tanwar, S., Gupta, R., ... & Sharma, R. (2022). A taxonomy of fake news classification techniques: Survey and implementation aspects. *IEEE Access*, *10*, 30367-30394.

Alameri, S. A., & Mohd, M. (2021, January). Comparison of fake news detection using machine learning and deep learning techniques. In *2021 3rd international cyber resilience conference (CRC)* (pp. 1-6). IEEE.

Abdullah, A., Awan, M., Shehzad, M., & Ashraf, M. (2020). Fake news classification bimodal using convolutional neural network and long short-term memory. *Int. J. Emerg. Technol. Learn*, *11*, 209-212.

Abdullah, A., Awan, M., Shehzad, M., & Ashraf, M. (2020). Fake news classification bimodal using convolutional neural network and long short-term memory. *Int. J. Emerg. Technol. Learn*, *11*, 209-212.

Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, *1*(1), 100007.

A. A. Nureni et al, "A Predictive Model for Benchmarking the Performance of Algorithms for Fake and Counterfeit News Classification in Global Networks," Sensors, vol. 24, (17), pp. 5817, 2024. Available: https://www.proquest.com/scholarly-journals/predictive-model-benchmarking-performance/docview/3104073906/se-2. DOI: https://doi.org/10.3390/s24175817.

E. Papageorgiou et al, "A Survey on the Use of Large Language Models (LLMs) in Fake News," Future Internet, vol. 16, (8), pp. 298, 2024. Available: https://www.proquest.com/scholarly-journals/survey-on-use-large-language-models-llms-fake/docview/3097908986/se-2. DOI: https://doi.org/10.3390/fi16080298.

T. V. Divya and F. Mohanty, "An Adaptive Word Vector Integrated With Enhanced Lstm For Classification Of Fake News," Journal of Electrical Systems, vol. 20, (7), pp. 4062-4071, 2024. Available: https://www.proquest.com/scholarly-journals/adaptive-word-vector-integrated-with-enhanced/docview/3081860750/se-2.

A. Sandu et al, "Numbers Do Not Lie: A Bibliometric Examination of Machine Learning Techniques in Fake News Research," Algorithms, vol. 17, (2), pp. 70, 2024. Available: https://www.proquest.com/scholarly-journals/numbers-do-not-lie-bibliometric-examination/docview/3048722652/se-2. DOI: https://doi.org/10.3390/a17020070.

S. Maham et al, "ANN: adversarial news net for robust fake news classification," Scientific Reports (Nature Publisher Group), vol. 14, (1), pp. 7897, 2024. Available: https://www.proquest.com/scholarly-journals/ann-adversarial-news-net-robust-fake/docview/3031478398/se-2. DOI: https://doi.org/10.1038/s41598-024-56567-4.

PDF, "MR-FNC: A Fake News Classification Model to Mitigate Racism," International Journal of Advanced Computer Science and Applications, vol. 15, (2), 2024. Available: https://www.proquest.com/scholarly-journals/mr-fnc-fake-news-classification-model-mitigate/docview/2992551137/se-2. DOI: https://doi.org/10.14569/IJACSA.2024.0150257.

S. V. Balshetwar, A. RS and D. J. R, "Fake news detection in social media based on sentiment analysis using classifier techniques," Multimedia Tools Appl, vol. 82, (23), pp. 35781-35811, 2023. Available: https://www.proquest.com/scholarly-journals/fake-news-detection-social-media-based-on/docview/2866560133/se-2. DOI: https://doi.org/10.1007/s11042-023-14883-3.

S. K and P. S. Thilagam, "Multi-layer perceptron based fake news classification using knowledge base triples," Appl. Intell., vol. 53, (6), pp. 6276-6287, 2023. Available: https://www.proquest.com/scholarly-journals/multi-layer-perceptron-based-fake-news/docview/2779961896/se-2. DOI: https://doi.org/10.1007/s10489-022-03627-9.

Y. Shen et al, "Fake News Detection on Social Networks: A Survey," Applied Sciences, vol. 13, (21), pp. 11877, 2023. Available: https://www.proquest.com/scholarly-journals/fake-news-detection-on-social-networks-survey/docview/2888111655/se-2. DOI: https://doi.org/10.3390/app132111877.

E. Shushkevich, M. Alexandrov and J. Cardiff, "Improving Multiclass Classification of Fake News Using BERT-Based Models and ChatGPT-Augmented Data," Inventions, vol. 8, (5), pp. 112, 2023. Available: https://www.proquest.com/scholarly-journals/improving-multiclass-classification-fake-news/docview/2882576357/se-2. DOI: https://doi.org/10.3390/inventions8050112.

Ş. E. Repede and R. Brad PhD., "A comparison of artificial intelligence models used for fake news detection," Bulletin of "Carol I" National Defense University, vol. 12, (1), pp. 114-131, 2023. Available: https://www.proquest.com/scholarly-journals/comparison-artificial-intelligence-models-used/docview/2877722002/se-2. DOI: https://doi.org/10.53477/2284-9378-23-10.

Y. Lu et al, "An Explainable Fake News Analysis Method with Stance Information," Electronics, vol. 12, (15), pp. 3367, 2023. Available: https://www.proquest.com/scholarly-journals/explainable-fake-news-analysis-method-with-stance/docview/2849001647/se-2. DOI: https://doi.org/10.3390/electronics12153367.

P. Moreno-Vallejo et al, "Fake News Classification Web Service for Spanish News by using Artificial Neural Networks," International Journal of Advanced Computer Science and Applications, vol. 14, (3), 2023. Available: https://www.proquest.com/scholarly-journals/fake-news-classification-web-service-spanish/docview/2807222530/se-2. DOI: https://doi.org/10.14569/IJACSA.2023.0140334.

D. Dementieva, M. Kuimov and A. Panchenko, "Multiverse: Multilingual Evidence for Fake News Detection," Journal of Imaging, vol. 9, (4), pp. 77, 2023. Available: https://www.proquest.com/scholarly-journals/multiverse-multilingual-evidence-fake-news/docview/2806552637/se-2. DOI: https://doi.org/10.3390/jimaging9040077.

E. Rahmanian, "Fake news: a classification proposal and a future research agenda," Spanish Journal of Marketing - ESIC, vol. 27, (1), pp. 60-78, 2023. Available: https://www.proquest.com/scholarly-journals/fake-news-classification-proposal-future-research/docview/2802184129/se-2. DOI: https://doi.org/10.1108/SJME-09-2021-0170.

K. Venkatachalam et al, "DeepFND: an ensemble-based deep learning approach for the optimization and improvement of fake news detection in digital platform," PeerJ Computer Science, 2023. Available: https://www.proquest.com/scholarly-journals/deepfnd-ensemble-based-deep-learning-approach/docview/2899137057/se-2. DOI: https://doi.org/10.7717/peerj-cs.1666.

K. A. Fawaz and A. K. Jawad, "Deep Dive into Fake News Detection: Feature-Centric Classification with Ensemble and Deep Learning Methods," Algorithms, vol. 16, (11), pp. 507, 2023. Available: https://www.proquest.com/scholarly-journals/deep-dive-into-fake-news-detection-feature/docview/2892938886/se-2. DOI: https://doi.org/10.3390/a16110507.

Y. Lu et al, "Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice," Systems, vol. 11, (9), pp. 458, 2023. Available: https://www.proquest.com/scholarly-journals/sustainable-development-information-dissemination/docview/2869639882/se-2. DOI: https://doi.org/10.3390/systems11090458.

P. L. Kemp, T. R. Alexander and C. N. Wahlheim, "Recalling fake news during real news corrections can impair or enhance memory updating: the role of recollection-based retrieval: Principles and Implications," Cognitive Research, vol. 7, (1), 2022. Available: https://www.proquest.com/scholarly-journals/recalling-fake-news-during-real-corrections-can/docview/2714991894/se-2. DOI: https://doi.org/10.1186/s41235-022-00434-1.

L. Chun-Ming et al, "Fake News Classification Based on Content Level Features," Applied Sciences, vol. 12, (3), pp. 1116, 2022. Available: https://www.proquest.com/scholarly-journals/fake-news-classification-based-on-content-level/docview/2636122630/se-2. DOI: https://doi.org/10.3390/app12031116.

P. L. Kemp et al, "Correcting fake news headlines after repeated exposure: memory and belief accuracy in younger and older adults: Principles and Implications," Cognitive Research, vol. 9, (1), pp. 55, 2024. Available: https://www.proquest.com/scholarly-journals/correcting-fake-news-headlines-after-repeated/docview/3096712807/se-2. DOI: https://doi.org/10.1186/s41235-024-00585-3.

Mohamad Rafad M. M. and P. Gopika, "Fake News Detection Using Machine Learning," International Research Journal of Innovations in Engineering and Technology, vol. 8, (2), pp. 138-142, 2024. Available: https://www.proquest.com/scholarly-journals/fake-news-detection-using-machine-learning/docview/3029998211/se-2. DOI: https://doi.org/10.47001/IRJIET/2024.802020.

PDF, "Twitter Truth: Advanced Multi-Model Embedding for Fake News Detection," International Journal of Advanced Computer Science and Applications, vol. 15, (8), 2024. Available: https://www.proquest.com/scholarly-journals/twitter-truth-advanced-multi-model-embedding-fake/docview/3108267341/se-2. DOI: https://doi.org/10.14569/IJACSA.2024.0150855.

M. Q. Alnabhan and P. Branco, "BERTGuard: Two-Tiered Multi-Domain Fake News Detection with Class Imbalance Mitigation," Big Data and Cognitive Computing, vol. 8, (8), pp. 93, 2024. Available: https://www.proquest.com/scholarly-journals/bertguard-two-tiered-multi-domain-fake-news/docview/3097829722/se-2. DOI: https://doi.org/10.3390/bdcc8080093.

W. Jin et al, "A veracity dissemination consistency-based few-shot fake news detection framework by synergizing adversarial and contrastive self-supervised learning," Scientific Reports (Nature Publisher Group), vol. 14, (1), pp. 19470, 2024. Available: https://www.proquest.com/scholarly-journals/veracity-dissemination-consistency-based-few-shot/docview/3095827568/se-2. DOI: https://doi.org/10.1038/s41598-024-70039-9.

PDF, "Cross-Modal Fine-Grained Interaction Fusion in Fake News Detection," International Journal of Advanced Computer Science and Applications, vol. 15, (5), 2024. Available: https://www.proquest.com/scholarly-journals/cross-modal-fine-grained-interaction-fusion-fake/docview/3072003155/se-2. DOI: https://doi.org/10.14569/IJACSA.2024.0150596.

J. Li et al, "Fake User Detection Based on Multi-Model Joint Representation," Information, vol. 15, (5), pp. 266, 2024. Available: https://www.proquest.com/scholarly-journals/fake-user-detection-based-on-multi-model-joint/docview/3059523222/se-2. DOI: https://doi.org/10.3390/info15050266.

W. Zhao et al, "Fake News Detection Based on Knowledge-Guided Semantic Analysis," Electronics, vol. 13, (2), pp. 259, 2024. Available: https://www.proquest.com/scholarly-journals/fake-news-detection-based-on-knowledge-guided/docview/2918725234/se-2. DOI: https://doi.org/10.3390/electronics13020259.

S. K. Uppada, P. Patel and S. B., "An image and text-based multimodal model for detecting fake news in OSN's," J Intell Inform Syst, vol. 61, (2), pp. 367-393, 2023. Available: https://www.proquest.com/scholarly-journals/image-text-based-multimodal-model-detecting-fake/docview/2880581580/se-2. DOI: https://doi.org/10.1007/s10844-022-00764-y.

J. Liu and S. Mi, "American literature news narration based on computer web technology," PLoS One, vol. 18, (10), 2023. Available: https://www.proquest.com/scholarly-journals/american-literature-news-narration-based-on/docview/2878351834/se-2. DOI: https://doi.org/10.1371/journal.pone.0292446.

I. Salah, K. Jouini and O. Korbaa, "On the use of text augmentation for stance and fake news detection," Journal of Information and Telecommunication, vol. 7, (3), pp. 359-375, 2023. Available: https://www.proquest.com/scholarly-journals/on-use-text-augmentation-stance-fake-news/docview/2842694193/se-2. DOI: https://doi.org/10.1080/24751839.2023.2198820.

B. Cui et al, "Intra-graph and Inter-graph joint information propagation network with third-order text graph tensor for fake news detection," Appl. Intell., vol. 53, (16), pp. 18971-18988, 2023. Available: https://www.proquest.com/scholarly-journals/intra-graph-inter-joint-information-propagation/docview/2858086086/se-2. DOI: https://doi.org/10.1007/s10489-023-04455-1.

B. Wang et al, "Multi-modal transformer using two-level visual features for fake news detection," Appl. Intell., vol. 53, (9), pp. 10429-10443, 2023. Available: https://www.proquest.com/scholarly-journals/multi-modal-transformer-using-two-level-visual/docview/2815842915/se-2. DOI: https://doi.org/10.1007/s10489-022-04055-5.

M. Berrondo-Otermin and A. Sarasa-Cabezuelo, "Application of Artificial Intelligence Techniques to Detect Fake News: A Review," Electronics, vol. 12, (24), pp. 5041, 2023. Available: https://www.proquest.com/scholarly-journals/application-artificial-intelligence-techniques/docview/2904838085/se-2. DOI: https://doi.org/10.3390/electronics12245041.

Ş. E. Repede, "Researching disinformation using artificial intelligence techniques: challenges," Bulletin of "Carol I" National Defense University, vol. 12, (2), pp. 69-85, 2023. Available: https://www.proquest.com/scholarly-journals/researching-disinformation-using-artificial/docview/2878460416/se-2.

K. Šteflovič and J. Kapusta, "Coreference Resolution for Improving Performance Measures of Classification Tasks," Applied Sciences, vol. 13, (16), pp. 9272, 2023. Available: https://www.proquest.com/scholarly-journals/coreference-resolution-improving-performance/docview/2856814019/se-2. DOI: https://doi.org/10.3390/app13169272.

S. Alyoubi, M. Kalkatawi and F. Abukhodair, "The Detection of Fake News in Arabic Tweets Using Deep Learning," Applied Sciences, vol. 13, (14), pp. 8209, 2023. Available: https://www.proquest.com/scholarly-journals/detection-fake-news-arabic-tweets-using-deep/docview/2842954229/se-2. DOI: https://doi.org/10.3390/app13148209.

M. Nasery, O. Turel and Y. Yuan, "Combating Fake News on Social Media: A Framework, Review, and Future Opportunities," Communications of the Association for Information Systems, vol. 53, pp. 833-876, 2023. Available: https://www.proquest.com/scholarly-journals/combating-fake-news-on-social-media-framework/docview/2834439858/se-2. DOI: https://doi.org/10.17705/1CAIS.05335.

M. A. Abdullah et al, "Web-Informed-Augmented Fake News Detection Model Using Stacked Layers of Convolutional Neural Network and Deep Autoencoder," Mathematics, vol. 11, (9), pp. 1992, 2023. Available: https://www.proquest.com/scholarly-journals/web-informed-augmented-fake-news-detection-model/docview/2812658322/se-2. DOI: https://doi.org/10.3390/math11091992.

C. Tang et al, "Long text feature extraction network with data augmentation," Appl. Intell., vol. 52, (15), pp. 17652-17667, 2022. Available: https://www.proquest.com/scholarly-journals/long-text-feature-extraction-network-with-data/docview/2737807043/se-2. DOI: https://doi.org/10.1007/s10489-022-03185-0.

M. Asutosh, T. Nithin and P. Prakasam, "Fake news detection and classification using hybrid BiLSTM and self-attention model," Multimedia Tools Appl, vol. 81, (13), pp. 18503-18519, 2022. Available: https://www.proquest.com/scholarly-journals/fake-news-detection-classification-using-hybrid/docview/2660203440/se-2. DOI: https://doi.org/10.1007/s11042-022-12764-9.

P. L. Kemp, V. M. Loaiza and C. N. Wahlheim, "Fake news reminders and veracity labels differentially benefit memory and belief accuracy for news headlines," Scientific Reports (Nature Publisher Group), vol. 12, (1), pp. 21829, 2022. Available: https://www.proquest.com/scholarly-journals/fake-news-reminders-veracity-labels/docview/2755191693/se-2. DOI: https://doi.org/10.1038/s41598-022-25649-6.

H. Alhakami et al, "Evaluating Intelligent Methods for Detecting COVID-19 Fake News on Social Media Platforms," Electronics, vol. 11, (15), pp. 2417, 2022. Available: https://www.proquest.com/scholarly-journals/evaluating-intelligent-methods-detecting-covid-19/docview/2700536496/se-2. DOI: https://doi.org/10.3390/electronics11152417.

Arianna D'Ulizia et al, "Fake news detection: a survey of evaluation datasets," PeerJ Computer Science, 2021. Available: https://www.proquest.com/scholarly-journals/fake-news-detection-survey-evaluation-datasets/docview/2543675877/se-2. DOI: https://doi.org/10.7717/peerj-cs.518.

Y. Jang et al, "Fake News Detection on Social Media: A Temporal-Based Approach," Computers, Materials, & Continua, vol. 69, (3), pp. 3563-3579, 2021. Available: https://www.proquest.com/scholarly-journals/fake-news-detection-on-social-media-temporal/docview/2568302814/se-2. DOI: https://doi.org/10.32604/cmc.2021.018901.