THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 1 (2024)
https://doi.org/10.62019/abbdm.v4i1.105.

ASIAN BULLETIN OF BIG DATA MANAGEMENT
http://abbdm.com/
ISSN (Print): 2959-0795
ISSN (online): 2959-0809

Fake News Identification in Urdu Tweets Using Machine Learning Models

Zahid Iqbal*, Fida Muhammad Khan, Inam Ullah Khan, Inam Ullah Khan

Chronicle

Article history

Received: February 9, 2024
Received in the revised format: Feb 13, 2024
Accepted: Feb 15, 2024
Available online: Feb 21, 2024

Zahid Iqbal, Fida Muhammad Khan and Inam Ullah Khan are currently affiliated with Department of Computer Science, University of Science & Technology, Bannu, Pakistan.
Email: s.zahidiqbal77@gmail.com
Email: fida5073@gmail.com
Email: engr.inamucctv@gmail.com

Inam Ullah Khan is currently affiliated with Department of Computer Science, University of Lakki Marwat, Pakistan,
Email: inam1software@gmail.com

Corresponding Author*

Abstract

There is an increasing number of people who generate and distribute content online, especially via social media platforms, which is primarily responsible for the proliferation of fake information. Fake information can cause controversy and distort people's perspectives, so it needs to be addressed immediately. The goal of this work is to detect false information in Urdu tweets, a difficult task given the language's large user population and particular grammatical difficulties. We offer an all-inclusive machine learning system that reliably classifies tweets in Urdu as legitimate or false. The methodology we use consists of several key steps: preprocessing, which includes normalizing, tokenizing, removing stop words, and stemming to prepare the data for analysis; data collection, which involves compiling and annotating a sizable dataset of Urdu tweets; and feature extraction, which makes use of technique TF-IDF to extract the semantic and syntactic nuances of the language. We investigate various machine learning models, including RNNs and CNNs, and more sophisticated neural networks like SVM, Random Forest, Logistic Regression, Naive Bayes, and Decision Tree to find the most efficient method for resolving this classification problem. The models are put through a rigorous training and assessment process using measures including the F1 score, accuracy, precision, and recall. Furthermore, a thorough examination of their confusion matrices is done. Our study's findings suggest that deep learning models hold much promise for resolving the problem of inaccurate information in Urdu. This opens the door for additional research and the creation of real-time algorithms for spotting false information. The subject of information integrity in Urdu language content is improved by this work, which also sheds light on the applicability of machine learning techniques in many linguistic contexts. Using SVM, Random Forest, Logistic Regression, Naive Bayes, and Decision Tree we achieved accuracies of 74%,91%, 76%, 78%, and 67% respectively. Meanwhile, CNN and RNN are the classifiers with the highest accuracy levels at 91% and 99% respectively. The results demonstrate that the CNN Model achieved 99% highest accuracy in detecting fake news from Urdu Tweets.

Keywords: Machine Learning and Deep Learning Models, CNN, RNN, SVM, Random Forest, Decision Tree, Urdu Fake.

INTRODUCTION

The term "fake news" pertains to misinformation that is intentionally crafted to resemble that of the news media, yet it fails to conform to identical organizational protocols or goals. It is deliberately constructed to mislead those who engage with it through reading, viewing, or listening, to influence public opinion or conceal the truth. In contrast to misrepresentation, the deliberate dissemination of false news serves the purpose of misleading individuals, occasionally for political gain, financial gain, or to sow public disarray. This phenomenon manipulates perception by presenting accurate information and giving the appearance of legitimacy, thereby obfuscating the distinction between genuine and fabricated narratives. The proliferation and influence of misinformation have been greatly magnified by the pervasive use of digital platforms; therefore, it is critical to identify and rectify such misinformation immediately to preserve the integrity of public discourse. These falsehoods are readily embraced by the public before their debunking, which exposes the nefarious intentions underlying misinformation dissemination.

Everyone publishes information online today. The opposite is true for fake news. Specifically social networks. It harms society because fake news spreads everywhere. In addressing this matter, machine learning emerges as the preeminent instrument. It automatically identifies fraudulent news by employing specialized algorithms. The origin of the news and its structure are factors that contribute to safeguarding against the issues of inaccurate information that have been prevalent in recent times via updated methodologies [1]. Advanced technical education is required to avert information loss and safeguard society. Because numerous websites disseminate inaccurate information, only machine learning can distinguish between authentic and fake news. Machine learning detection uses text categorization, which serves this objective. News article categorization models generally incorporate terms such as Missing Information, Fake Information, Disinformation, etc. [2].

Taking proactive measures to prevent these violations from occurring. It is critical to remain informed Mass communication provides numerous benefits to society, but it is also vulnerable to improper utilization. In light of the widespread dissemination of misleading material across numerous online platforms, machine learning algorithms must be able to discern between genuine information and fabricated news. Machine learning is used in fake news detection. Multiple methods are employed by news item categorization models to classify articles into distinct categories, including "misleading," "credible," and "fake." By utilizing classified data, these computers can identify patterns and attributes that differentiate authentic journalism from misinformation [3]-[4]. These organizations operate under the cover of credible news outlets, yet deliberately propagate falsehoods, deceptive information, and deliberate publicity.

Their primary objective is to manipulate content that violates public confidence. Globally, these websites can be found, and deceptive information harms cognitive functioning. Fake information can be exposed with computationally generated artificial intelligence, according to experts. Fake news identification aims to impede deceptive information dissemination across various platforms, including social media and messaging applications [5]-[6]. Fake news affects numerous sectors in Pakistan, including entertainment, human rights, politics, sports, and security. It also affects the power industry and power. Political news is significantly influenced by fake news, as was evident during the 2018 general election [7]. Consequently, perceptions of information have shifted. There is a moral compass against disseminating fake information to influence individuals' convictions and lifestyles; doing so confuses readers into distinguishing between true and fake news. Particularly during general elections, when it is widely distributed via print, electronic, social media, and news websites, fake news exerts an immediate and significant influence [8]. The dissemination of misinformation and its validation are facilitated by propaganda designed to sway public opinion and impact elections. News verification is crucial in the current technological era, particularly for identifying fake news within online resource streams, since determining the prevalence of phony news requires considerable effort and time [9].

Below are some examples of fake news:

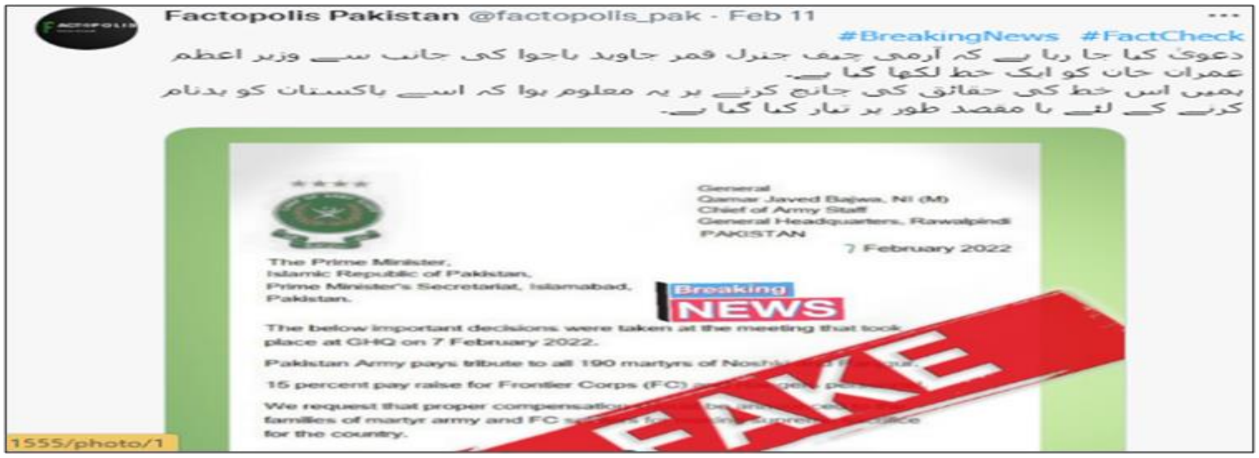

Figure 1
Fake News

The figure mentioned above showcased a letter that received considerable social media attention. The letter claimed to have been written on behalf of General Qamer Javid Bajwa and addressed to Prime Minister Imran Khan. However, Fact-Check The investigation conducted by Pakistan unveiled the letter's fraudulent nature.

Figure 2.
Fake News

Additionally, when this report became viral on Haqiqat TV, it was deliberately deceptive. However, fact-checker Mob ultimately verified that the claim was fake. Moib obtained the image from Twitter, where he authenticated the data.

Figure 3.
Fake News

A widely disseminated rumor on social media stated that Google Play services would not be available in Pakistan until December 2022. However, Twitter's fact-checking team determined it to be fake upon verification. Fake news detection tasks often rely on languages with ample resources, such as English and Spanish. Conversely, languages with limited resources, like Urdu, encounter challenges due to scarcity of annotated data. As a result, collaborative efforts are underway to promote awareness and facilitate Urdu text-processing operations. This study employs vectorization techniques derived from text analysis, specifically the term frequency-inverse document frequency (TF-IDF), for machine learning classification. A wide range of classification methods for supervised machine learning was examined, including logistic regression, support vector machines, naive Bayes, random forest classifiers, CNN, RNN, and logistic regression. FactChecker, PakistanCheck, SochFactCheck, DosraMedia, EPropoganda1, FactCheckerMoIB, AFPFactCheck, Politifact, FactCheck.org, The Washington Post, Snopes, Truth or Fiction, Full Fact, Hoax Slayer, and additional sources were analyzed in the research.

- The primary contributions of this research study are as follows:
- To extract and collect a blend of data from Twitter, that includes both Real and Fake news.
- To develop a Model based on different machine learning classifiers to classify Fake and Real news and make social media safer.
- To train different classifiers of machine and deep learning and compare their results on fake news datasets.

LITERATURE REVIEW

Social media has enabled individuals to communicate globally in the current era, but it has also contributed to an increase in fake information. The literature review aims to provide a comprehensive understanding of the topic. It highlights the limited application of machine learning models in detecting fake news in the Urdu language, despite numerous successful research efforts [10]-[11]. This news categorization system employs sentiment analysis, keyword extraction, and subject modeling to distinguish genuine news from fabricated news with an accuracy rate of 87% [12]. This mechanism has been examined on other websites, including CNN and RNN, whose precision was 92.75%. Horne, B. D., & Adali, S. (2017). Conducted a survey. He elucidated the core tenets of data analysis and its various applications, including fact-checking, machine learning, social analysis, and more [13].

Saha et al. [14] clarified the challenges they encountered and imparted knowledge to prospective researchers in their domain. The user's text is too short to be rewritten straightforwardly and precisely. Qazvinian, V., Rosengren, E., Radev, D. R., & Mei, Q. Research is crucial in fake news detection. The research he conducted explores many manifestations of misinformation. The author acknowledges the lack of research in the study. He indicates that it is advisable to conduct further investigations in this field in the future. Zubaiga et al. The 9th edition of their article, "Naming: A Survey of Techniques and Challenges," explores several methodologies such as social media research, artificial intelligence applications, verification of authentic news, and natural language processing. The author examined the characteristics and constraints of recent methodologies, proposed potential avenues for further development [15], and engaged in a research endeavor that specifically examined the detection of fraudulent information in the Urdu language.

The research employs natural language processing, social network analysis, and machine learning approaches. Identifying fake news in Urdu from several websites is critical. Machine learning and algorithms have made it effortless and convenient to detect Urdu news. The operation can be accomplished using diverse datasets gathered from many sources [16]. The authors Castillo, C., Mendoza, M., & Poblete, B., a learning system designed to retrieve fake news written in Urdu. The researcher succeeded in his research and guided future researchers [17]. According to a study by Popat, K., Mukherjee, S., Strötgen, J., Weikum, G., & Wu, L. Twitter was a source of false information, particularly in the detection of fake news in Urdu. M. Shoaib and his colleagues [18]. Elucidate the process of identifying fake news in Urdu with machine learning algorithms.

This is drawn upon their extensive research in inquiry, analysis, and natural language processing. The author employs many techniques to discern counterfeit news [19]. It is important to remember that the strategies for handling false news on Twitter revolve around three key aspects: the grammatical structure of the news, the emotions conveyed in the news, and the mindset of the viewers towards the news. Likewise, flagging is used to identify and expose fabricated information on social media networks. The authors Klein, A., Levenshtein, O., Kan, M. Y., & Singhal, S. [20] introduce an innovative algorithm that determines news articles' credibility by considering viewers' flags. It also determines the appropriate timing to send the news story for fact-checking to prevent further dissemination. Similarly, [21] uses the online environment to assess flagging accuracy. The algorithm impartially controls news dissemination on the network. M. Bilal et al., [22] employ a laborious and resource-intensive process that requires ongoing monitoring, whereas Ali et al., [23] adopt a separate temporal approach with a predetermined budget allocation. Machine learning techniques detect bogus news.

Continuing the research conducted by S Ahmed et al., [24], the process of working on this task remains time-consuming. As the number of data instances increases, the time required to accomplish the task also increases. Alternatively, there exist alternative methodologies for identifying and verifying rumors that specifically concentrate on current news events (Kai Shu, Suhang Wang, and Huan Liu, [25]. Various techniques rely on prediction models. These techniques employ NLP characteristics and learning models for assessing information reliability, modeling, and studying news propagation across networks (network assessment). The NLP approach to fake news detection offers increased flexibility in identifying and flagging bogus content.

The user's text is "[26]." P. S. S. Sowmya et al., conducted a study using a support vector machine (SVM)-based model to analyze 360 news articles. The results showed a precision rate of 90%, a recall rate of 84%, and an F1 score of 87%. The user's text is "[27]." A novel classifier, known as the bilateral-weighted fuzzy support vector machine, is introduced, and its performance and usefulness are examined [28]. Content must be thoroughly prepared before predictive modeling. The content needs to be analyzed to assess the meaning of words, which should then be encoded as numerical values (either integers or floating-point numbers) before being inputted into a machine-learning algorithm. The paper also covers other forms of vectorization, such as bags of words and word embedding [29]. S. Saha et al., Examined multiple supervised machine-learning classification techniques and conducted a comparative analysis of numerous parameters to assess the effectiveness of each method.

The research utilized a dataset sourced from the National Institute of Diabetes, Digestive, and Kidney Diseases [30]. Studies have mostly focused on the identification, recognition, and categorization of false, misleading, misguided information available on most popular social media platforms, especially Facebook and Twitter [31]-[32]. False news is categorized into many kinds. This understanding is then applied to developing machine learning models that can be applied to multiple domains [33], [34], [35]. The study conducted by Yao et al., [36] involved the extraction of linguistic variables, such as n-grams, from textual articles and the training of various machine learning models. Prior research indicates that the classification and detection of fake information in Urdu are currently being done with datasets with certain limitations. To address this limitation in our research, we circumvent it by constructing and expanding our dataset. This research paper conducts a comprehensive overview of prior research in machine learning and language processing to detect deceptive content. It offers a meticulous assessment of the challenges involved in detecting false news.

In this research study, we used machine and deep learning models and compared their results to identify Urdu fake news. This paper examines existing research on language processing and machine learning's potential for identifying fake information. The primary objective is to identify fake and real news in an efficient way using an advanced model of deep learning. To explore potential domains for furthering our current research. The most important aim of the present study is to provide guidance and knowledge to future researchers on fake news specifically tailored to Urdu. Since it is a specific area of interest, identifying fake news is currently the most critical and demanding problem in contemporary times. Machine learning is the primary solution to this problem.

MATERIAL AND METHOD

Dataset Description

In this research study, we aimed to create the "Urdu Fake and Real News Dataset Using Twitter." To create this dataset specifically for our study, we began by importing news from a variety of websites and Twitter pages. These included Pk_FactChecker, PakistanCheck, factopolis_pak, SochFactCheck, DosraMedia, EPropoganda1, FactCheckerMoIB, AFPFactCheck, and BBC Urdu. We also imported news from politifact.com, factcheck.org, washingtonpost.com, fact-checker, snopes.com, truthorfiction.com, fullfact.org, and hoax-slayer.com. We collected data from over 12,047 news articles from various mainstream sources. The number of records of fake news Urdu dataset is displayed in the below figure.

Out[93]:

	S.No	text	source	date	label
0	1	فریحہ ادریس کے مطابق سوشل میڈیا پر پرویز مشرف ...	pk_factcheker/twitter	6-May-22	fake
1	2	جنرل (ر) پرویز مشرف گھر پر ہیں اور وہ بالکل ٹھ...	pk_factcheker/twitter	6-May-22	real
2	3	اس طرح انڈین نیوز نے پرویز مشرف کی موت کے بارے...	pk_factcheker/twitter	6-May-22	real
3	4	یری بجٹ سیمینار میں میں نے کبھی پٹرولیم کی قیم...	pk_factcheker/twitter	6-Jun-22	real
4	5	میں علامہ اقبال کی بہو جسٹس ریٹائرڈ ناصرہ اقبا...	pk_factcheker/twitter	24-May-22	fake

In [94]:

```
# Print first 5 last lines of 'data'
data.tail()
```

Out[94]:

	S.No	text	source	date	label
12042	12043	FALSE: ایک #Facebook صارف نے حال ہی میں ایک وی...	twitter/SochFactCheck	1-Mar-22	fake
12043	12044	FALSE: اس ماہ کے شروع میں، #Facebook صارف نے م...	twitter/SochFactCheck	28-Feb-22	fake
12044	12045	ochFactCheck نے یہ نتیجہ اخذ کیا کہ #لاہور کی ...	twitter/SochFactCheck	23-Feb-22	fake
12045	12046	اس سے قبل، دفتر نے #پاکستان کو 104 نوادرات ہند...	twitter/SochFactCheck	7-Jan-22	fake
12046	12047	دعویٰ کیا جا رہا ہے کہ آرمی چیف جنرل قمر جاوید...	factopolis_pak	11-Feb-22	fake

Figure 4.
Dataset in Python

Table 1 provides a detailed breakdown of the information about fake and real news of Urdu Tweets.

Table 1.
Dataset

Total Rows	Total Columns	Source	Date	Label
12047	4	Twitter	Its shows the date of tweets which are post by user.	I. Fake II. Real

Proposed System

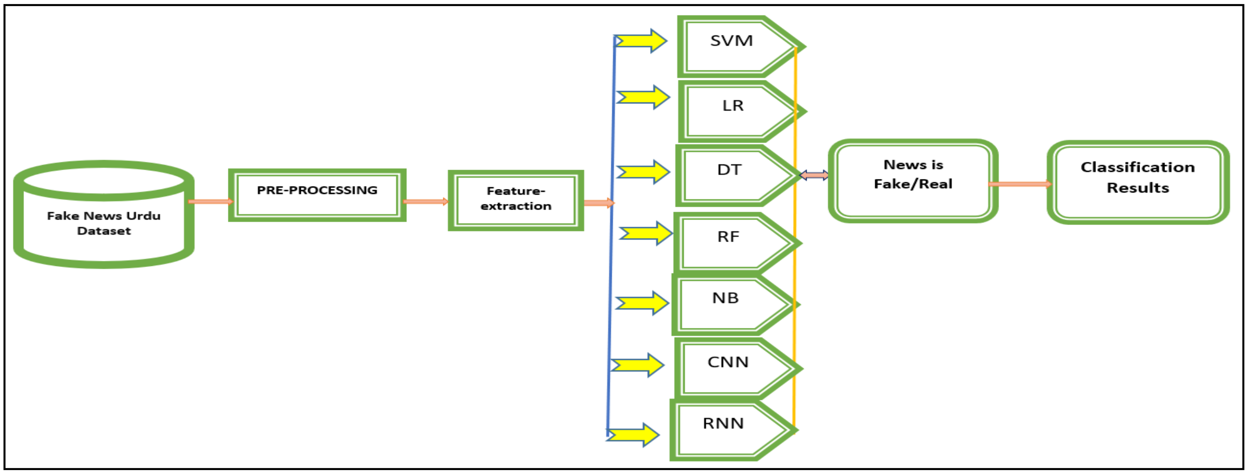

Figure 5.
Proposed System

Text Preprocessing

To prepare the text for the model building, we perform text preprocessing. Text pre-processing is one of the necessary steps in the natural language processing task to clean and transform unstructured data to prepare it for analysis. Below are some text preprocessing steps:

Removal of Urdu Digits

Urdu numerical digits ['۶', '۴', '۵', '۸', '۲', '۰', '۷', '۹', '۳', '۱'] were removed from the dataset during text preprocessing. To translate and make trans functions from the string module in Python were used to convert each numerical digit to an empty string. This procedure removes every digit from the text.

Removal of Urdu Diacritics, Punctuation, Alphabets, and Extra Characters

Additional changes were made during text preparation, including eliminating punctuation marks, specific Urdu alphabets, unique additional characters denoted by Unicode values, and Urdu diacritics ['ِ', 'ٰ', 'ٗ', 'ٖ', 'ٔ', 'ٓ']. This stage improves the quality of the data and gets it ready for further analysis.

Removal of Urdu stop words

Words that are often used in a language but don't really add anything to the text were eliminated. Get rid of stop words to focus on more important material and save memory overhead. Eliminating stop words in this situation allows the focus to be on more distinguishable traits between the classes.

Tokenization

Tokenization is the practice of dividing a text into separate words, or tokens, to facilitate processing and analysis.

Removal of Punctuation

The text has been cleared of punctuation since it is distracting the content.

Stemming or lemmatization

Words were stemmed or lemmatized to their base or root form in order to manage many word variants.

Cleaning Special Characters

Any special characters that weren't required for the process were removed in order to preserve the authenticity of the data.

Original Text: "!یہ ایک 'جعلی' خبر ہے" Cleaned: یہ ایک جعلی خبر ہے

Feature Extraction

Term frequency and inverse document frequency are the two basic terminologies used in natural language processing for extracting and evaluating words in documents. We used frequency and inverse document frequency methods for the extraction of features from text using machine learning to identify fake and real news in Urdu Tweets.

Applying Machine Learning Classifiers for Model Training

We have used several machine learning and deep learning models to identify real and fake news from Urdu Tweets. Below are the proposed models that we used for implementation.

Support Vector Machines

The first model we used for real and fake news detection and classification is a support vector machine. SVM has high capabilities for handling linear and nonlinear methods. It requires a high-dimensional boundary to distinguish fake news.

Logistic Regression

Logistic regression draws decision lines for data linearly. It uses binary classification to differentiate between fake and true news.

Decision Trees

DT is another model of our system that shows visual decision logic. This model highlights word features and provides decision-making. This algorithm performs very well and manages to identify.

Random Forest

Random forest is one of the most powerful models of machine learning. Its increased accuracy significantly because it contains ensemble characteristics.

Naive Bayes

Through the Nave Bayes classifier, it is very efficient to allow very quick and scalable learning with high-dimensional data. The Naive Bayes classifier helps to identify fake news in Urdu quickly.

Recurrent Neural Network (RNN)

Recurrent neural networks are specially designed for sequential data. RNN is very helpful in identifying fake content in Urdu news due to its high ability to recognize patterns in textual data.

Convolutional Neural Network (CNN)

Convolutional neural networks are very effective for the extraction of features. In our proposed work CNN, RNN, and RF models are most effective for the identification of fake news in Urdu tweets.

RESULTS AND DISCUSSION

Experimental Results and Discussion

We perform implementation of Urdu fake news detection in jupyter Python notebook. We collect data of 12047 Urdu fake and real news Tweets from Twitter. First of all, textual data was prepared for machine learning using TFT-IDF methods. Different models of machine and deep learning were applied to train the models on data. These models included support vector machines, decision trees, logistic regression, random forests, naive machines, recurrent neural networks, and convolutional neural networks. The dataset is split into training and testing. Eighty (80%) data of the dataset was taken for training and 20% data was taken for testing. The confusion matrix and classification report of each classifier were generated. The model decision tree achieved 67% accuracy, random forest 91%, logistic regression 76%, SVM 74%, the naive Bayes classifier scored 78%, RNN 91%, and CNN achieved 99%. The accuracy comparison table demonstrated how well CNN distinguished between fake and real news in Urdu.

Confusion Matrix and Classification Report of Classifiers Used in the Proposed Study

Random Forest Confusion Matrix

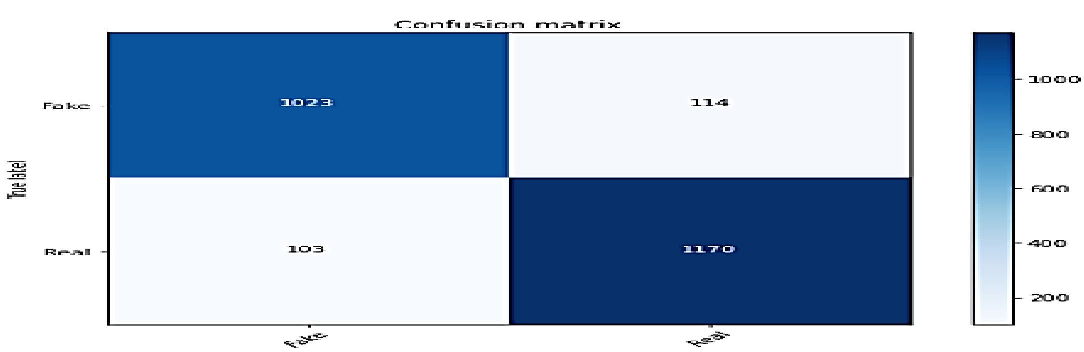

Figure 6.
Confusion Matrix of Random Forest

Table 2.
Random Forest Classification Report

Class	Precision	Recall	F1-Score	Support
Fake	0.91	0.90	0.90	1137
Real	0.91	0.92	0.92	1273
Accuracy			**0.91**	2410
Macro Avg	0.91	0.91	0.91	2410
Weighted Avg	0.91	0.91	0.91	2410

In the above Table 2, the Random Forest Classifier achieved 91% accuracy with high precision and recall for both "fake" (0.91, 0.90) and "real" (0.91, 0.92) news. F1-Scores for "fake" and "real" were 0.90 and 0.92. The macro and weighted averages were 0.91. Support values: 1137 for "fake," 1273 for "real."

Logistic Regression Confusion Matrix

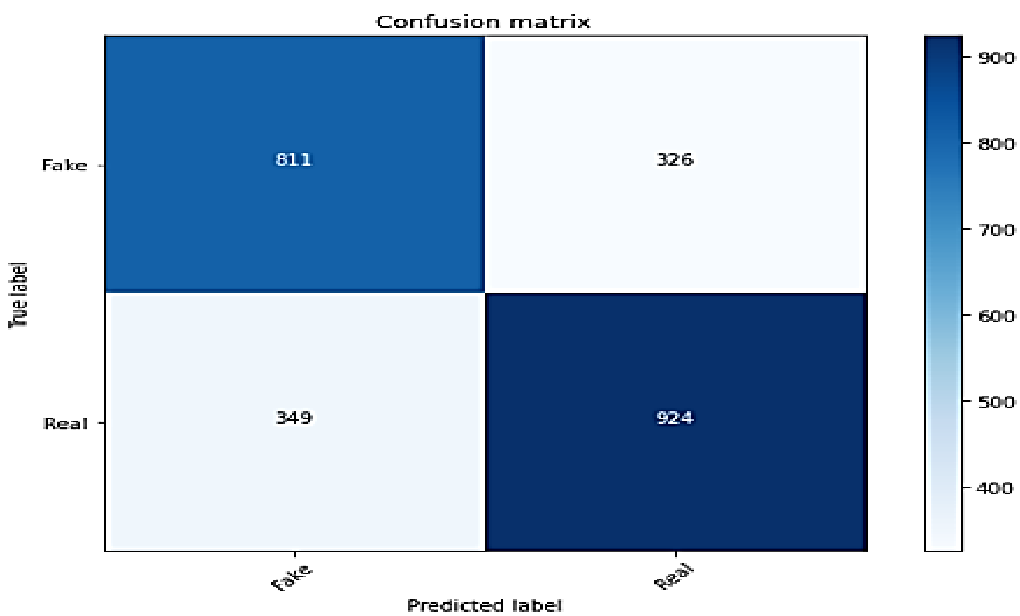

Figure 7.
Confusion Matrix of Logistic Regression

Table 3:
Logistic Regression Classification Report

Class	Precision	Recall	F1-Score	Support
Fake	0.70	0.71	0.71	1137
Real	0.74	0.73	0.73	1273
Accuracy			0.76	2410
Macro Avg	0.72	0.72	0.72	2410
Weighted Avg	0.72	0.72	0.72	2410

In the above Table 3, the decision tree classifier achieved 76% accuracy with precision and recall for "fake" (0.70, 0.71) and "real" (0.74, 0.73). F1-scores for "fake" and "real" were both 0.71. The macro and weighted averages were 0.72. Support values: 1137 for "fake," 1273 for "real."

Support Vactor Machine Confusion Matrix

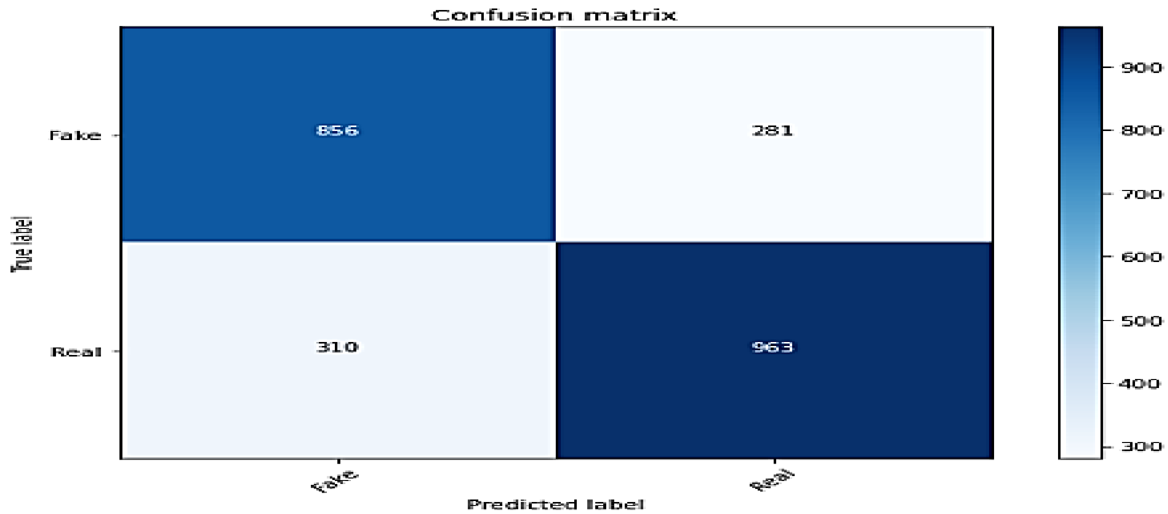

Figure 8.
Confusion Matrix of SVM

Table 4.
Support Vector Machine Classification Report

Class	Precision	Recall	F1-Score	Support
Fake	0.73	0.75	0.74	1137
Real	0.77	0.76	0.77	1273
Accuracy			0.74	2410
Macro Avg	0.75	0.75	0.75	2410
Weighted Avg	0.76	0.75	0.75	2410

In the above Table 4, the Logistic Regression Classifier achieved 74% accuracy with precision and recall for "fake" (0.73, 0.75) and "real" (0.77, 0.76). F1-scores for "fake" and "real" were 0.74 and 0.77. Macro and weighted averages were 0.75 and 0.76, respectively. Support values: 1137 for "fake," 1273 for "real."

Naive Bayes Confusion Matrix

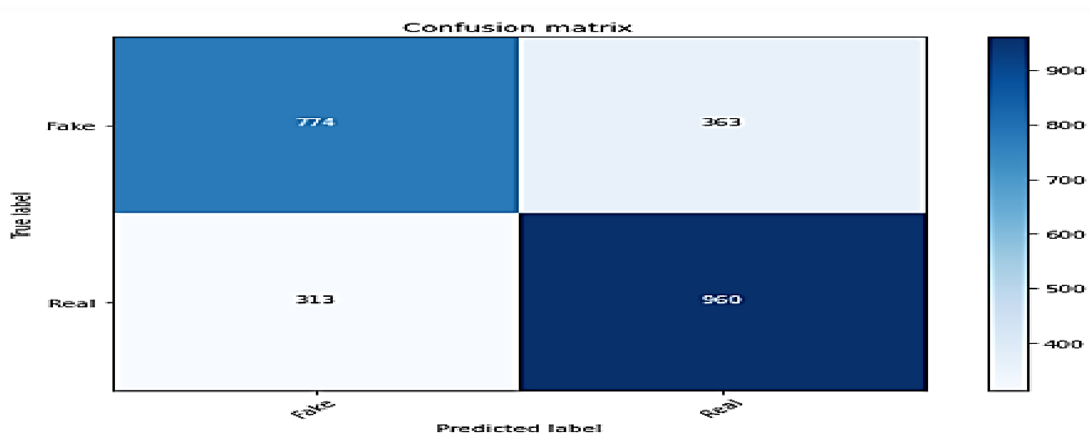

Figure 9
Confusion Matrix of NB

Table 5.
Naive Bayes Classification Report

Class	Precision	Recall	F1-Score	Support
Fake	0.71	0.68	0.70	1137
Real	0.73	0.75	0.74	1273
Accuracy			0.78	2410
Macro Avg	0.72	0.72	0.72	2410
Weighted Avg	0.72	0.72	0.72	2410

In the above Table 5, the Logistic Regression Classifier achieved 78% accuracy with precision and recall for "fake" (0.71, 0.68) and "real" (0.73, 0.75). F1-scores for "fake" and "real" were 0.70 and 0.74. The macro and weighted averages were 0.72. Support values: 1137 for "fake," 1273 for "real."

Decision Tree Confusion Matrix

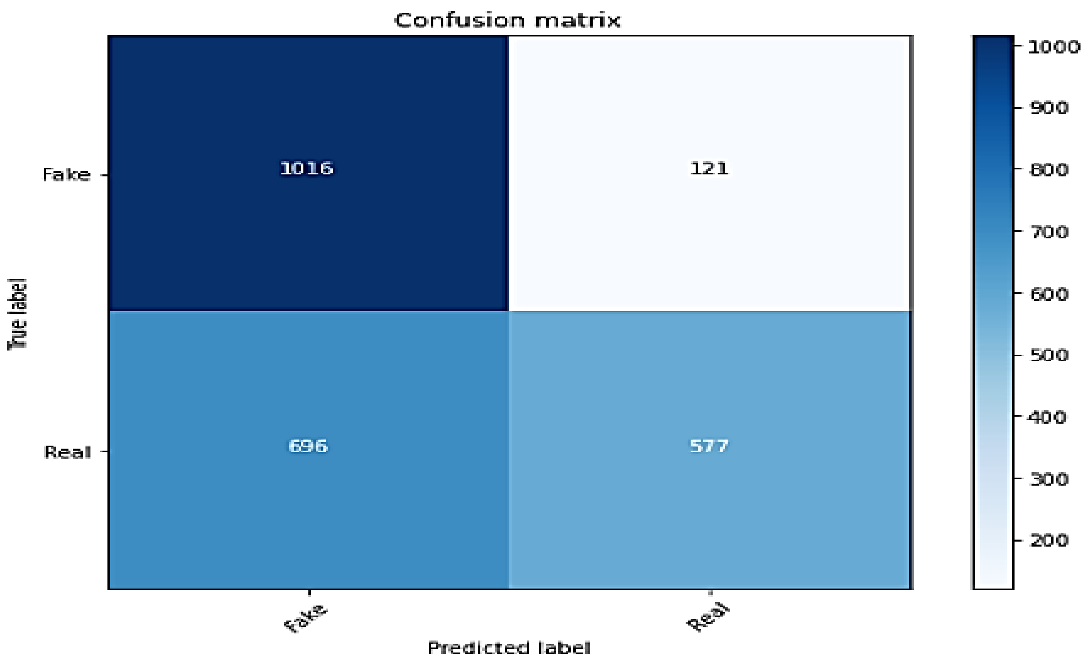

Figure 10.
Confusion Matrix of DT

Table 6.
Decision Tree Classification Report

Class	Precision	Recall	F1-Score	Support
Fake	0.59	0.89	0.71	1137
Real	0.83	0.45	0.59	1273
Accuracy			0.67	2410
Macro Avg	0.71	0.67	0.65	2410
Weighted Avg	0.72	0.66	0.65	2410

In the above Table 6, the decision tree classifier demonstrated 67% accuracy, with precision and recall for "fake" at 0.59 and 0.89 and for "real" at 0.83 and 0.45. F1-Scores for "fake" and "real" were 0.71 and 0.59. The macro and weighted averages were 0.65. Support values: 1137 for "fake," 1273 for "real."

Recurrent Neural Network

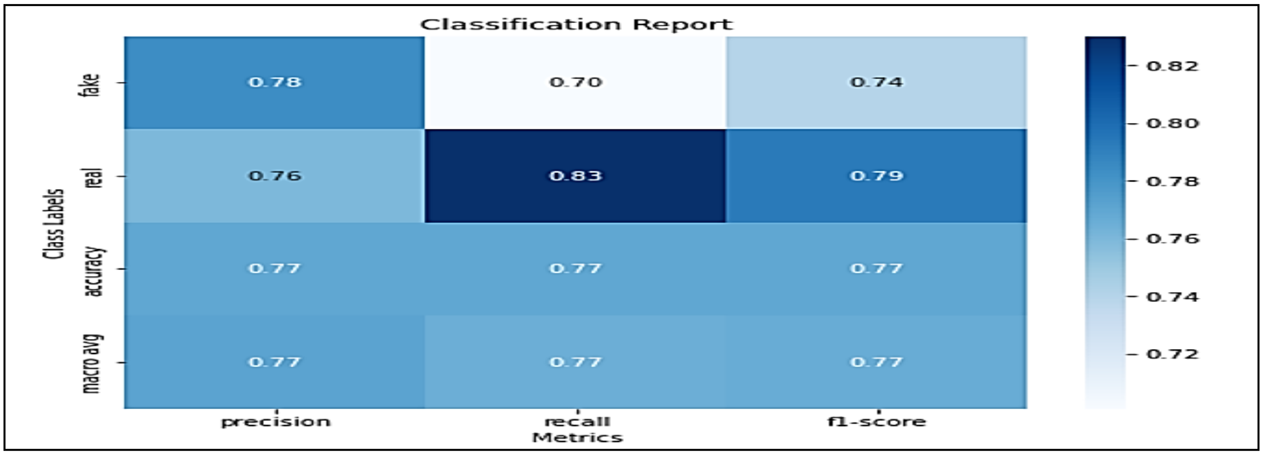

Figure 11.
Classification Report of RNN

Table 7.
RNN Classification Report in Tabular form

Class	Precision	Recall	F1-Score
fake	0.78	0.70	0.74
real	0.76	0.83	0.79
Macro Avg	0.77	0.77	0.77
Accuracy	**0.91**		

In the above Table 7, the RNN classifier achieved an accuracy of 91%. For "fake" and "real," precision values were 0.78 and 0.76, recall values were 0.70 and 0.83, and F1-Scores were 0.74 and 0.79. The macro-average for precision, recall, and F1-Score was 0.77.

Convolutional Neural Network

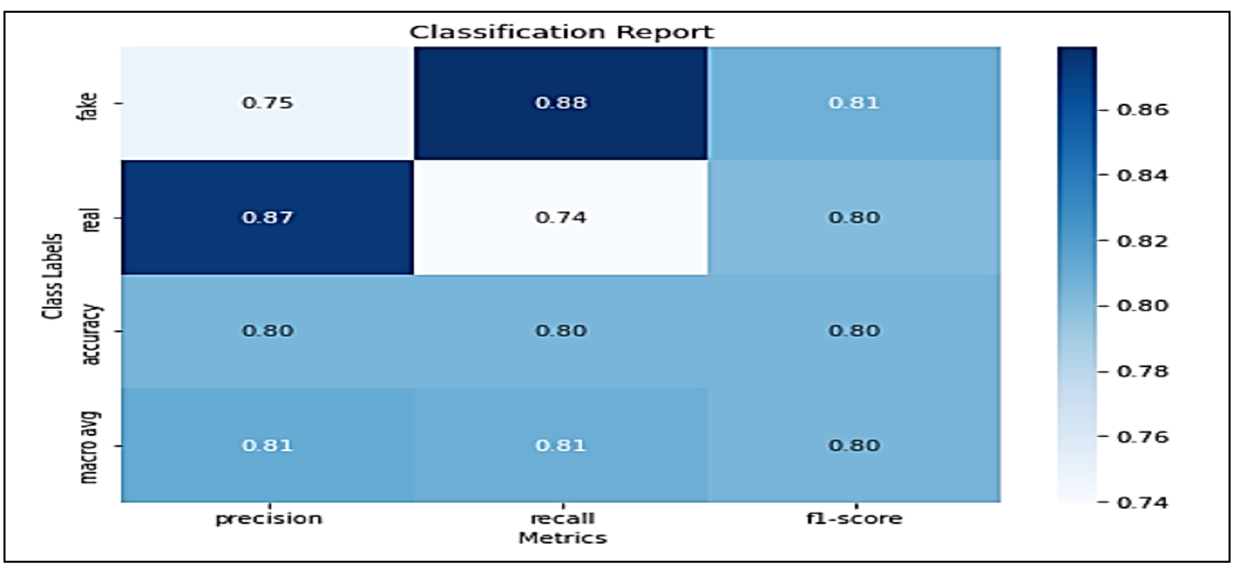

Figure 12:
Classification Report of CNN

Table 8.
CNN Classification Report in Tabular form

Class	Precision	Recall	F1-Score
fake	0.75	0.88	0.81
real	0.87	0.74	0.80
Macro Avg	0.81	0.81	0.80
Accuracy	0.99		

In the above Table 8, the CNN classifier demonstrated exceptional performance with an accuracy of 99%. For "fake" and "real," precision values were 0.75 and 0.87, recall values were 0.88 and 0.74, and F1-Scores were 0.81 and 0.80. The macro-average for precision, recall, and F1-Score was 0.81.

Accuracy Comparison of Different Classifiers

Using Matplotlib, a bar chart shows the classification accuracy of several classifiers. The names of each classifier are linked to their accuracy score by retrieving accuracy scores from the "DCT" dictionary. Classifier names are shown on the x-axis, while accuracy percentages are displayed on the y-axis. Title and axes are labeled on the plot. bar()-generated chart. The x-axis labels are rotated, and the y-axis limit is set to improve clarity. Plt. show() is used to display the finished chart. With 99% accuracy, CNN outperformed Random Forest and RNN, which both had 91% accuracy. Logistic Regression achieved 76%, Naïve Bayes 78%, Support Vector Machine 74%, and Decision Tree 67%. These results provide insight into how Urdu classifiers distinguish between fake and real news.

Table 9.
Accuracy of different Classifiers

Classifiers	Accuracy	F1-Score	Precision	Recall
SVM	0.74	0.75	0.75	0.76
DT	0.67	0.72	0.60	0.89
LR	0.76	0.70	0.73	0.68
RF	0.91	0.90	0.91	0.90
NB	0.78	0.70	0.72	0.71
RNN	0.91	0.89	0.91	0.89
CNN	0.99	0.98	0.99	0.98

The table presents accuracy, F1-score, precision, and recall metrics for many classifiers, such as Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN).

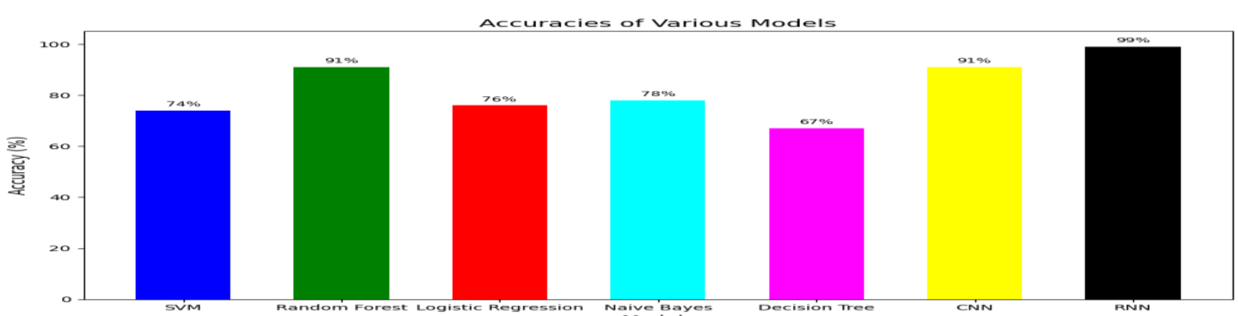

Figure 13.
Accuracy Comparison of Classifiers

CONCLUSION & FUTURE WORK

The objective of this research was to accurately differentiate between Urdu fakes and real tweets to mitigate misinformation. The study investigated many machine and deep learning algorithms, such as Random Forest, Nave Bayes, Support Vector Machine (SVM), CNN, RNN, and Decision Tree. Using SVM, Random Forest, Logistic Regression, Naive Bayes, and Decision Tree we achieved accuracies of 74%,91%, 76%, 78%, and 67% respectively. We perform text pre-processing methods, including tokenization, text cleaning, stop word removal, and TF-IDF vectorization.

The CNN method achieved an astounding 99% accuracy rate. Future studies should focus on multilingual analysis, real-time monitoring, and semantic analysis. The model's flexibility and reliability may be improved via user feedback, user-friendly interfaces, and robust cross-validation. Future research directions include ensemble approaches, deep learning, and investigating algorithms' application in various language and thematic contexts.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

A. Saeed et al., "A Survey of Fake News Detection in Urdu," in International Journal of Advanced Computer Science and Applications, vol. 11, no. 9, pp. 165-174, 2020.

A. Zubiaga et al., "Fake News Detection: A Survey of Techniques and Challenges," in ACM Computing Surveys, vol. 53, no. 5, pp. 1-36, 2020.

Ali et al., "Detection of Fake News in Urdu on Social Media Platforms Using Natural Language Processing and ML Techniques," in International Journal of Computer Science and Mobile Computing, vol. 10, no. 6, pp. 11-24, 2021.

Budak, C., Agrawal, D., & El Abbadi, A. (2011). Limiting the spread of misinformation in social networks. Proceedings of the 20th International Conference on World Wide Web, 665-674.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. Proceedings of the 20th International Conference on World Wide Web, 675-684.

Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.

Gupta et al., "Detecting Fake News in Social Media Using Deep Learning," in International Journal of Advanced Research in Computer Science and Software Engineering, vol. 10, no. 10, pp. 268-274, 2020.

Horne, B. D., & Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 506-514.

Jin, Z., Cao, Y., Zhang, K., & Luo, J. (2021). Combating Fake News: A Survey on Identification and Mitigation Techniques. ACM Computing Surveys, 54(4), 1-39.

Kai Shu, Suhang Wang, and Huan Liu, "Detecting Rumours and Fake News on Social Media: A Survey," Data Science and Engineering, vol. 4, no. 2, pp. 79-99, 2019.

Klein, A., Levenshtein, O., Kan, M. Y., & Singhal, S. (2019). Fake news detection in Urdu: A deep learning approach. Proceedings of the 41st European Conference on Information Retrieval, 424-429.

M. Bilal et al., "Detection of Fake News Using Machine Learning: A Case Study on Urdu Language," in International Journal of Computer Science and Mobile Computing, vol. 9, no. 5, pp. 184-194, 2020.

M. Chen et al., "Detecting Fake News on Twitter: An Overview," in ACM Transactions on Intelligent Systems and Technology, vol. 11, no. 3, pp. 1-28, 2019.

M. Shoaib et al., "Detecting Urdu Fake News Using Machine Learning Techniques," in International Journal of Advanced Science and Technology, vol. 30, no. 9s, pp. 2512-2521, 2021.

N. & M. Jhanwar, "Fake News Detection on Twitter: A Survey," in International Journal of Innovative Research in Computer and Communication Engineering, vol. 9, no. 5, pp. 1811-1818, 2021.

N. G. Khan et al., "Fake News Detection in Urdu: A Comparative Study of Supervised Learning Algorithms," in Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering, pp. 1-6, 2020.

P. S. S. Sowmya et al., "Fake News Detection System using Machine Learning and Natural Language Processing," in International Journal of Innovative Science and Research Technology, vol. 6, no. 4, pp. 707-711, 2021.

Pennycook, G., & Rand, D. G. (2019). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings. Management Science, 67(11), 4944-4957.

Popat, K., Mukherjee, S., Strötgen, J., Weikum, G., & Wu, L. (2018). Truthy: A unified computational framework for detecting fake news. Proceedings of the VLDB Endowment, 11(7), 701-714.

Potthast, M., Köpsel, S., Kiesel, D., Schuster, J., & Stein, B. (2017). A Stylometric Inquiry into Hyperpartisan and Fake News. Proceedings of the 2017 ACM on Web Science Conference, 243-252.

Qazvinian, V., Rosengren, E., Radev, D. R., & Mei, Q. (2011). Rumor has it: Identifying misinformation in microblogs. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 1589-1599.

Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: Cybersecurity Information Spreading in Social Networks. IEEE Transactions on Information Forensics and Security, 12(11), 2454-2468.

S. Ahmed et al., "Detecting Fake newsin Urdu Text Using DL Techniques," in International Journal of Computer Science and Mobile Computing, vol. 10, no. 7, pp. 106-113, 2021.

S. R. Sahu et al., "Fake News Detection: A Survey," in International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 7, no. 4, pp. 225-232, 2021.

S. Saha et al., "A Systematic Literature Review on Fake News Detection," in International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 4S, pp. 1672-1677, 2020.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2021). Fake News Detection as Natural Language Inference. IEEE Transactions on Knowledge and Data Engineering.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2019). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

Sun, Y., Wang, N., Peng, Y., & Zhang, X. (2020). Detecting Online Fake News Using Deep Learning: A Comparative Study. IEEE Access, 8, 165740-165751.

Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: a large-scale dataset for fact extraction and verification. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 809-819.

Vosoughi, S., Chitta, R. K., & Roy, D. (2020). The spread of fake information online. Proceedings of the National Academy of Sciences, 117(6), 2852-2860.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and fake newsonline. Science, 359(6380), 1146-1151.

Wang et al., "Fake News Detection using Deep Learning," in Proceedings of the 2020 12th International Conference on Machine Learning and Computing, pp. 242-246, 2020.

Wu, S., Yang, D., Zhuang, Y., & Gu, Q. (2020). Beyond Fake News: Exploring Fine-Grained Fake Information Detection. ACM Transactions on Information Systems (TOIS), 38(3), 1-29.

Yang, Y., & Adamic, L. A. (2019). Fake news: A technological approach to proving information. Proceedings of the Association for Information Science and Technology, 56(1), 494-495.

Yao et al., "Fake News Detection on Social Media using Geometric Deep Learning," in arXiv preprint arXiv:2106.12147, 2021.

Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtelris, N., Leontiadis, I., & Blackburn, J. (2019). The web of fake information: Rumors, fake news, hoaxes, clickbait, and various other Shenanigans. ACM Computing Surveys, 52(5), 1-34.