A Data-Driven Framework for Fake News Detection Via Web Scraping and Machine Learning Approach

Sarvesh Kumar[1]; Dr. Yusuf Perwej[2]; Farheen Siddiqui[3]; Ankit Shukla[4]; Dr. Nikhat Akhtar[5]

[1;3;4;5]Assistant Professor, [2]Professor
[1]Department of Computer Science & Engineering,
Shri Ramswaroop Memorial University, Deva Road, Lucknow
[2]Department of Computer Science & Engineering,
Shri Ramswaroop Memorial University, Deva Road, Lucknow
[3]Department of Computer Science & Engineering,
Shri Ramswaroop Memorial University, Deva Road, Lucknow
[4]Department of Computer Science & Engineering,
Shri Ramswaroop Memorial University, Deva Road, Lucknow
[5]Department of Computer Science & Engineering,
Goel Institute of Technology & Management, Lucknow

Publication Date: 2025/06/21

Abstract: A great deal of misinformation has been circulated on a global scale in recent years due to the explosion of social media. The spread of false information has been worsened by recent political events. Some 1835 news stories were completely made up, like the one about "Bat-men on the moon." There has to be a system in place for checking claims, particularly those that get a lot of attention before being debunked by reliable sources. In order to properly categorize and identify fake news, a plethora of machine learning techniques have been used. The technique for spotting fake news inside datasets is the focus of this study. Online traditional news stories and news from other sources make up the bulk of the collection. The outcomes are compared to those of deep learning and traditional machine learning methods applied to the datasets, as well as long short-term memory (LSTM). Several example procedures are compared with the recommended methodology, and the results are given. In a number of respects, our work is superior than current methods. This approach has laid the groundwork for a system that can spot several red flags associated with fake news, classifying the material as either genuine or fraudulent and making decisions easier.

Keywords: *Fake Profile, Web Scraping, Natural Language Processing (NLP), Detection, Fake News, Data Mining, LIAR Dataset, Machine Learning.*

How to Cite: Sarvesh Kumar; Dr. Yusuf Perwej; Farheen Siddiqui; Ankit Shukla; Dr. Nikhat Akhtar (2025) A Data-Driven Framework for Fake News Detection Via Web Scraping and Machine Learning Approach. *International Journal of Innovative Science and Research Technology*, 10(6), 1391-1404. https://doi.org/10.38124/ijisrt/25jun1003

I. INTRODUCTION

People, organizations, and businesses alike are feeling the effects of the fast-growing problem of fake news. Manipulation of public opinion, harm to reputations, drops in stock market values, and other risks to global health are some of the serious and relevant outcomes that could emerge from this major issue in today's linked and modern world. The rapid growth of internet deceit has rendered manual verification of fake news impractical due to its complexity, labor intensity, and lack of transparency. What we call "fake news" is reporting that doesn't really exist but makes it seem like it does. False, misleading, erroneous, manufactured, altered, or satirical claims, as well as fictitious connections and parodies, may be included. Thus, false information may significantly affect several parts of life [3]. There are a lot of methods that have been proposed to detect fake news. There has been much research on many different topics related to feature extraction, representation, classification, and model construction. Because fake news may have long-lasting

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

effects, spotting it has become increasingly challenging. It developed from Cold War-era deception, which may have its roots in propaganda from the 17th century. In recent years, this problem has been further worsened by the rise of social media platforms. In recent years, social media sites like Instagram, Twitter, and Facebook have emerged as hubs for the instantaneous sharing and retrieval of information. A large body of research shows that in industrialized nations, social media accounts for about 50% of news consumption [7]. There are a lot of methods that have been proposed to detect fake news. The importance of social media cannot be denied, and it has shown to be a useful tool during crises, as seen by its role in spreading news as it happens. However, one drawback of social media's accessibility is the quick dissemination of false information.

Users of social media platforms are able to alter content by inserting their own views and prejudices, in contrast to more conventional forms of media like print or television. This could have far-reaching effects on the meaning or context of the news. Someone on social media may put the financial or ethical well-being of another person, group, or society at risk by creating or manipulating news stories, adding their own biased views or beliefs, and changing the way the story is viewed. Sarcasm, comedy, misleading advertisements, false political claims, and baseless rumors are all components of misleading news. When disinformation affects a community or government, members look to reputable news outlets for answers. However, human fact-checking is impossible due to the large amount of information that is either faked or uses machine-generated news. Due to the possibility of inaccurate information about source or authorship attribution, the aesthetic distinctions from human-generated work are not always obvious. A number of technical obstacles stand in the way of addressing the pressing societal issue of disinformation and its correction. A number of well-known companies have decided to deal with these issues. Google has created a library of fake movies to aid academics in their quest to identify them, while Facebook and Microsoft have launched the DeepFake Detection Challenge.

II. BACKGROUND

Because of the unique, fluid, quick, and portable nature of the material provided by social media platforms, people rely heavily on them for knowledge, information, and current events. Their day-to-day lives are greatly affected by the news's content. It might change the way they feel, think, or see things. The 2016 US presidential election is the best example of this phenomena. A great chance to change one's outlook and make the most of the circumstance has presented itself. On rare occasions, news outlets may distort the truth to suit their own agendas. Modern deception and skilled misinformation pose a threat to reality. Intentional misinformation, attention seeking, swaying public opinion, or reputational harm are common goals of these misleading articles. Over the last ten years, several methods have been developed to address the issue of identifying disingenuous news reports. Logistic regression, decision trees, k-nearest neighbours, random forests, and support vector machines (SVMs) are all part of the ensemble technique that Jiang et al. [15] provide. When tested on real-time news, all of these methods achieved an impressive 85% accuracy rate. The public, private, and governmental sectors all utilize these concepts as a basis for their news broadcasts. To distinguish between trustworthy microblogs and rumours based on user behaviour, Chen et al. [17] developed an unsupervised learning method that uses auto-encoders and recurrent neural networks [18]. The analysis found that the suggested model was accurate 92.49% of the time and had an F1 score of 89.16%. Negative effects of disinformation on communities and governments To clarify or answer to disputes about the validity of a false claim, use other reputable news sources [19]. The sheer volume of deliberately false or exploitative content makes manual fact-checking impractical at times. Since deceit may be linked to acknowledging sources or writers, it may not always be possible to tell HGM from false information based on visual style alone.

In order to detect misleading material, Xu et al. [21] used domain credibility and content comprehension. They have depended on the number of domain registrations, the news's longevity, and the domain's popularity. One Arabic corpus for research on credibility classification was created by Al Zaatari et al. [22]. Using quantitative methods, Allcott et al. [24] examined the impact of social media misinformation on American voters in the 2016 U.S. Presidential General Election. Using the BuzzFeed dataset, the authors compared bogus news URLs that were either legitimate or manipulative. When it comes to spreading false information, posts with images are shared almost eleven times more often than those without [25]. As a result, false information often includes visual components, and created pictures may frequently be fascinating and emotional. Therefore, it's important to link these emotional reactions with the qualities of an image [27].

False photos are fascinating and spectacular because of this, and visual material is frequent in disinformation. Therefore, it is crucial to associate these mental cues with visual characteristics of an image [27]. When Shu et al. (2017) look at how social and psychological theories relate to fake news, they find that similar groups tend to accept erroneous information as reality. This happens because it is our nature to seek for, absorb, and trust data that supports our current worldview. There has been a deluge of new research on deceptive news in the last year, outnumbering the papers reviewed in [28]. In order to achieve erroneous user identification, Elhadad et al. [29] use decision tree models, dual-path deep semi-supervised learning, and deep neural networks. In situations when real-time disinformation detection is required, these models are quite useful for producing quick judgments, despite their modest performance. The influence of the twenty most popular articles on Twitter, as measured by the number of retweets, was examined by Potthast et al. [30]. For this task, five annotators were recruited using the web-based crowdsourcing platform Amazon Mechanical Turk. From various angles, the study was assessed by a variety of questionnaires. Automated fake news detection research takes four factors into account: the reliability of the source,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

the style of writing or content-based analysis, the patterns of social transmission, and the veracity of the information. We have graphically represented these opinions and related components by examining a number of research articles [33].

Research by Neves et al. [34] delves into methods for identifying social cognition, blogging networks, and multimedia disinformation. These methods are well-suited for use as the backbone of a large-scale system to detect fake news since they can accurately identify category-specific false news with a 90% success rate. Works that have shown to be pertinent to rumours make up the literature reviewed in [35]. The two tasks of rumour classification and fake news detection have many commonalities in terms of characteristics and approaches. We provide a comprehensive overview of pertinent remedies from similar scenarios to remedy the shortcoming of this research. Zhou et al. [37] investigated the potential of social media to compile the viewpoints of a large user base. To extract event-invariant features, Wang et al. suggested EANN, which consists of a multimodal feature extractor, a false news detector, and an event discriminator. This study uses adversarial networks to improve generalizability and gather data that is independent of events. The VGG19 network that has already been trained and is utilized to get visual representations; this network is mentioned in Jin et al. [39]. The authors have explored machine learning approaches to enhance rumour identification in their next study. In order to improve these frameworks, they have examined the difficulties of rumour distribution, rumour categorization, and dishonesty. In a study by Wang et al., a new dataset was presented for the purpose of detecting fake news [40]. Some have proposed a mixed strategy to combat disinformation. To obtain representations of meta-data, they used a convolutional neural network (CNN) followed by an LSTM (long short-term memory) neural network) [41]. In addition to being complicated and in need of several adjustments, their suggested model has a test set error rate of 27.4 percent, suggesting poor performance. In addition, a comparison strategy for detecting fake rumors has been developed, as stated by Yang et al. [42]. During the English riots of 2011, researchers looked at how narrative upgrades based on false rumors could have helped. Despite the fact that their research into the 2013 Boston Marathon bombings produced several interesting news items, the majority of them were assumptions that had a substantial influence on the stock market [44].

III. FAKE NEWS

Academic literature classifies several types of deceptive or damaging information as follows: hoax, misleading news, conspiracy theories, conspiracies, parody, satirical news, propaganda, disinformation, misinformation, hoax, misleading news, rumor, clickbait, and so on. The six main types of academic uses of the term—satire, parody, fabrication [47], manipulation, advertisement, and propaganda—are outlined by Lim et al. [46]. It is possible to further categorize the six groups according to the degree of factual correctness and the degree of purpose to mislead. A multi-classifier approach to combating fake news was proposed by Thorne et al. [48]. It used a multilayer perceptron (MLP) based on ReLU activation with word2vec to handle headline processing and tf-idf to handle article body processing. Additionally, it averaged word2vec for headlines and article body, and tf-idf for bigrams and unigrams in article body. A multilayer perceptron and dropout were then used for processing the article body and headlines, followed by logistic regression with L2 regularization and word2vec concatenation. The use of tf-idf alone will not be sufficient to identify fake news.

Fig 1 The Example of Fake News

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

The way news is displayed on homepages and news feeds causes users to interact solely with specific sorts of material [49]. When people band together, their views tend to become more polarized since they share similar ideals. There are two main reasons why consumers are naive and easily misled. Two biases are shown by users: naïve realism, which is when people like to believe news stories that align with their own thoughts or views (based on rationalism or the Theory of Perception), and confirmation bias, when people choose to believe things that support their own opinions. Misinformation is spread about eleven times more often in postings that include images than in ones that do not [51].As a result, misleading information often incorporates visual components, and inaccurate pictures frequently capture attention and cause anxiety [38]. Associating the picture's qualities with these emotional responses is crucial. Unlike conventional object-level descriptions, these behavioral features are exclusive to outward appearance [52]. Misleading spectators may happen with either real, unmodified photos or digitally changed, phony ones. utilizing pictures from an earlier event to portray a more recent one is an example of image misappropriation or utilizing images out of context [54]. As a result, regular picture datasets are not a good fit for this false image classification problem. Posts and profiles on social media that are inaccurate help disseminate false information. From time to time, reputable news outlets portray this as factual [55]. When anything may be reported as news, the difference between truth and fiction is blurry. False information has been around for a while. At this moment, how important is this issue? The main cause of this is that false information can be easily created, shared, and absorbed by our never-ending news cycle. The instances in figure 1 also show that fake news is harder to spot.

IV. CHALLENGES ASSOCIATED WITH MISINFORMATION

Disinformation tactics rely on the internet and social media, yet they vary greatly between platforms. The tools and services that manipulate and spread content across pertinent social media platforms are essential to the spread of disinformation. From the most basic (such purchasing likes and followers) to the most creative, there is no shortage of social media tools and services available today. There are service providers that allegedly manipulate internet surveys and others who force website owners to erase content. There are a plethora of resources available for underground and mainstream social media marketing [61]. Recent events have highlighted the "fake news" issue as a major threat to responsible media and well-informed public discourse. Democracy is based on the tenets of mutual trust, faith in institutions, and the veracity of information. When hostile non-state actors or foreign governments launch propaganda operations, it may taint the information environment, cutting off public discourse and weakening confidence [62]. In the early months of 2017, the Fake News challenge was launched. Second, the social media site is an essential part. A social platform is necessary to access these services and technologies, which might be used inappropriately to spread false information. Their part in spreading disinformation is vital, especially because people are spending more time on these sites to keep up with the news and information. There are several ways in which the creation and distribution of false information pose serious threats to the safety of the country.

Therefore, identifying false news is a vital aim for improving the reliability of information shared on online social networks. Various academics have used various algorithms, techniques, tools, and tactics to detect fake news on social media platforms throughout the years [64]. The fourth and most crucial aspect is motive, which sheds light on the real goal of the misinformation campaign or fake news. Sometimes, the only motivation is the hope of making money via advertising. In other cases, the goals could be anything from purely political to downright illegal. No amount of well-intentioned disinformation can change the reality of the situation until it changes anything. We discovered a dearth of literature on the topic of developing web- or mobile-based technological solutions to alert people to the dangers of misleading news [66]. For example, think about the debates surrounding the 2016 US presidential election. There was a lot of animosity in the campaign conversations. Civility and faith in the nation's basic institutions have declined, according to poll respondents, since the election, as opposing ideological factions have hardened their stances. Almost 70% of respondents felt that civility had decreased, and fewer than 30% trusted media corporations, according to the poll.

V. LIAR DATASET

Misinformation is a major problem in modern culture, even yet people rely on the internet and online for important information every day. The use of supervised learning [69] is one approach to identify false news by assessing the veracity of the assertion based on linguistic patterns, posture data, and other criteria. There has been a significant expansion in the body of data pertaining to the detection of fake news [68]. However, most studies don't evaluate claims based on supporting evidence, context, or external validation. We have a larger and more comprehensive collection of bogus news stories than any other dataset out there. Our collection includes photographs and text as well as metadata and statistics on discussion. The LIAR Dataset, which includes more than 12.8K samples from various forms of fake news (as seen in figure 1), is an advanced multimodal dataset that we wholeheartedly recommend. For the purpose of identifying false news, the LIAR dataset is accessible to the public. We gathered 12.8K concise comments from POLITIFACT.COM under various settings; they were all hand-labeled over a decade. An in-depth analytical study and links to the original articles are both available on the website. Validation of study findings is another potential application of this dataset. Compared to the largest previously published datasets of this sort, this new dataset contains considerably more false news stories. A POLITIFACT.COM editor has verified the accuracy of the 12.8K concise human-validated assertions that make up the LIAR dataset. The corpus statistics are shown in Table 1.

 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

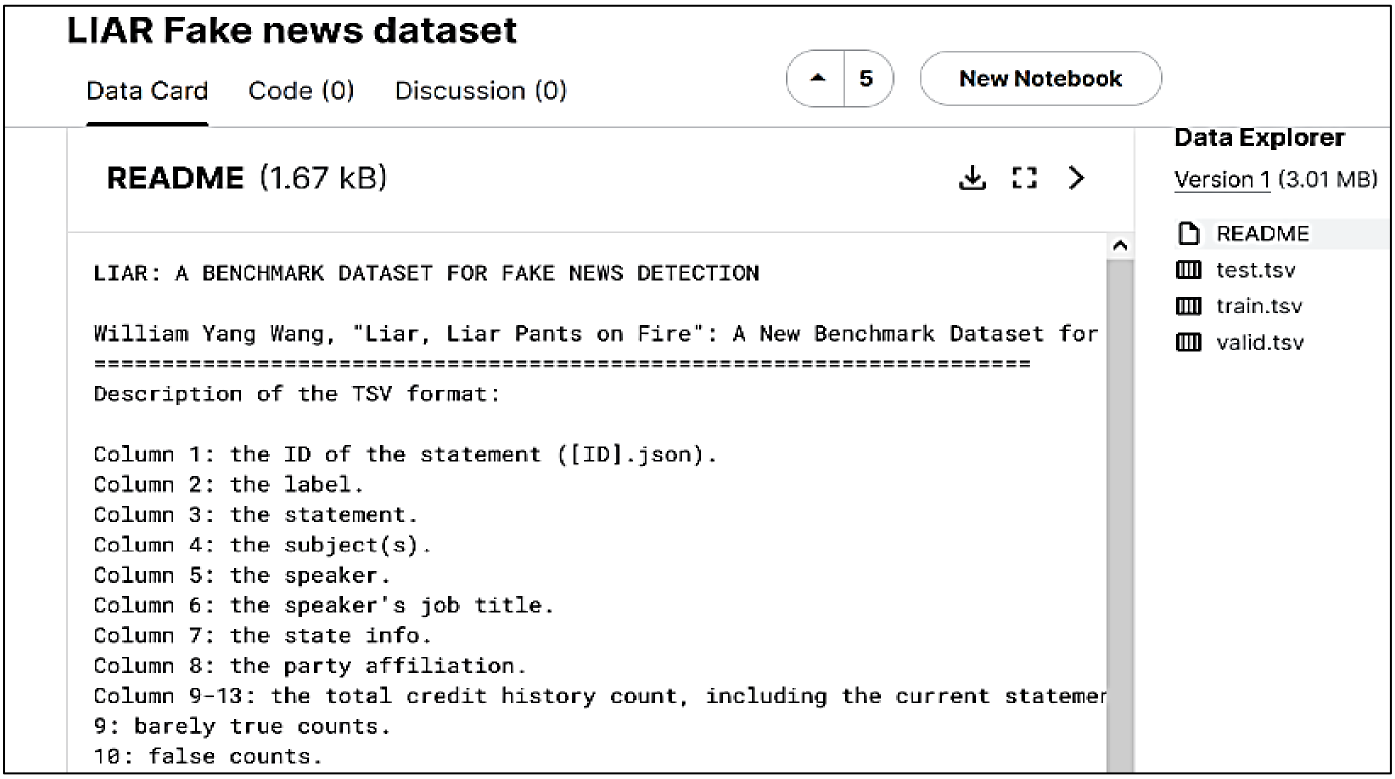

Fig 2 The Fake News LIAR Dataset

We merged the full-flop, half-flop, and no-flip labels into their corresponding false, half-true, and true labels after an initial analysis to eliminate duplicate labels. There are six subtle categories that we consider for the honesty evaluations: pants-fire, false, barely-true, half-true, mostly-true, and true. With the exception of 1,050 pants-fire incidents, the LIAR dataset displays a very even distribution of labels, with occurrences ranging from 2,063 to 2,638 for every other label. This study aims to analyze 200 randomly selected cases by looking at their detailed analytical reports and related judgments. There is a nearly equal split between Democrats and Republicans in the LIAR dataset, and many of the postings come from social media platforms. Each speaker's broad variety of meta-data includes their party memberships, current employment, home states, and credit histories, among many other things. For every single speaker, there is a record of their bogus claims in their credit history [70].

Table 1 Statistics for the LIAR Dataset

Dataset Statistics	**Set Size**
Training	10,269
Validation	1,284
Testing set	1,283
Avg. statement length (tokens)	17.9
Top-3 Speaker Affiliations	Set Size
Democrats	4,150
Republicans	5,687
None (e.g., FB posts)	2,185

VI. MACHINE LEARNING

Machine learning is a subfield of artificial intelligence. Data and sophisticated algorithms are the building blocks of machine learning, which uses examples drawn from past data to address current issues. The machine takes on more human traits as a result of its learning capabilities. Surprisingly, machine learning is already making waves in a number of industries. An astounding number of machine learning applications have emerged to automate procedures and solve problems in a variety of sectors. This is mostly attributable to developments in processing power, improvements in machine learning techniques, and the accessibility of supplemental data. Countless complex and modern network management and operational problems have undoubtedly found a solution in machine learning. A great deal of machine learning research has focused on niche networking technologies or niche networking companies. The ability to infer and filter data is made possible by machine learning. Learning entails doing more than just taking in data; it also necessitates putting that data to use and improving it over time. Finding and using hidden patterns in "training" data is the main goal of machine learning. Applying the found patterns allows for the classification or alignment of new data with existing categories [74]. The primary emphasis of our research is on algorithms that categorize data as true or fake news. Supervised learning is where most of the methods that were examined fit.

➢ *Long Short-Term Memory (LSTM)*

In sequence prediction challenges, LSTM recurrent neural networks may learn order dependencies. In theory, a Long Short-Term Memory (LSTM) recurrent unit would reject unnecessary input while attempting to remember all

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

previously received information by the network. This is accomplished by using many "gates" activation function levels, each of which serves a unique purpose. Every LSTM recurrent unit stores a vector called the Internal Cell State, which, in theory, defines the information that the previous LSTM recurrent unit maintained. A challenging subfield of deep learning is LSTMs, or long short-term memory networks. It may be challenging to understand LSTMs and how to apply concepts like bidirectional and sequence-to-sequence in the domain.

➢ *Logistic Regression*

Logistic regression is the suitable regression strategy to use when the dependent variable is dichotomous (binary). Similar to other regression studies, logistic regression serves as a predictive analysis [75]. Logistic regression elucidates the association between a binary dependent variable and one or more independent factors. Logistic regression may be categorized into three primary categories according on the characteristics of the dependent variable.

➢ *Random Forest*

The supervised learning method referred to as Random Forest is used for classification and regression tasks. However, classification concerns are mainly resolved. A forest consists of trees, and those with a greater density of trees often exhibit enhanced health. The random forest method constructs decision trees from data samples, derives predictions from each tree, and use voting to ascertain the optimal choice. The ensemble approach averages the results and mitigates overfitting, making it preferable to an individual decision tree.

➢ *Decision Tree (DT)*

The decision tree is one of the most efficient supervised learning techniques for classification and regression tasks. A tree structure akin to a flowchart is established by depicting each internal node as a test on an attribute, each branch as a test outcome, and each leaf node as a class designation. The termination criterion, such as the maximum tree depth or the minimum sample size required for node division, is attained when the training data is systematically partitioned into subsets based on attribute values. This algorithm is among the most effective. Moreover, Random Forest, a very effective machine learning method, use it to train on diverse subsets of training data.

➢ *Convolutional Neural Network (CNN)*

Convolutional neural networks (CNNs) are a category of deep learning methodologies adept in image processing and recognition. This architecture consists of many layers, including convolutional layers, pooling layers, and fully linked layers. Convolutional neural networks (CNN), which use grid-like matrices for feature extraction from datasets, were derived from artificial neural networks (ANN) [78]. The crucial element of a CNN is its convolutional layers, which use filters to extract features from the input picture, such as edges, textures, and forms. The output from the convolutional layers is then sent to pooling layers to down-sample the feature maps and retain the most significant information. The output from the pooling layers is then used in one or more fully connected layers to forecast or classify the picture. Convolutional layers are often succeeded by activation layers, pooling layers, and hidden layers in convolutional neural networks (CNNs) [79].

➢ *Recurrent Neural Network (RNN)*

One kind of neural network that uses the outputs from one stage as inputs for the next stage is called a recurrent neural network (RNN). In traditional neural networks, the inputs and outputs are completely separate entities. On the other hand, remembering what came before is necessary for guessing what will come next in a phrase [80]. To address this, a Recurrent Neural Network (RNN) with a Hidden Layer was created. The most important and distinctive aspect of an RNN is its hidden state, which stores sequence information. The memory state is another name for this state, which refers to the network's previous input. With consistent settings for each input, it generates the output by applying the same operation to all inputs or hidden layers. When contrasted with other neural networks, this simplifies the parameter set.

VII. WEB SCRAPING

The process of automatically retrieving data from a webpage is called web scraping. To get specific data from a website, one may use a web scraper tool to scan the page, analyze the data, and then extract it. The data is saved in a structured format that can be easily used in spreadsheets or apps, such Excel, JSON, or XML. The copied URL triggers a server request when run with the open web scraper code [81]. After receiving our request, the server sends the data and makes the HTML or XML page available to you. After that, the code finds the data, pulls it out, and then parses the XML or HTML file. In order to efficiently gather and validate data from several websites and social media platforms, our technique makes use of our unique code. The purpose of web scraping is to retrieve data. Gathering unstructured data and transforming it into a more useful format is the goal.

Fig 3 The Web Scraping Model

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

VIII. THE SUGGESTED APPROACH

By proposing a system that can detect and remove fake news from search engine results and social media news feeds, the idea tackles problems related to disinformation. The goal is to use AI to identify statements and publications that could include biased or inaccurate information. Text extraction and matching, which involves searching different data sources and comparing them with the input news; named entity recognition, which produces token-level, fine-grained outputs; document and entity-level sentiment analysis, which examines sentences for polarity, emotions, negations, sarcasm, tone, and bias; document classification, and stance classification are among the many tasks involved in the challenge of detecting fake news [83]. Downloading and integrating the approach into the end user's browser or news feed application is possible. All input texts and feeds must be managed by performing the required pre-processing activities. It begins by reading the validation, training, and testing datasets and performing processing operations like stemming and tokenization. Next, we carry out the exploratory data analysis, which involves examining the distribution of the response variable and evaluating the data quality, including any missing or null values. Linguists use a technique called lemmatization to classify a word's inflected forms and examine them as a whole, or lemma. As an example, the words "run," "runs," "ran," and "running" all belong to the same group of words that are related by inflection. It is a system for assigning each piece of text or piece of data a distinct identity. The following follows from this: as seen in figure 4, the reasoning provided will be applied to each word as the smallest item when a text is tokenized. The smallest item (or token) to be processed would be each phrase if we tend to tokenize sentences.

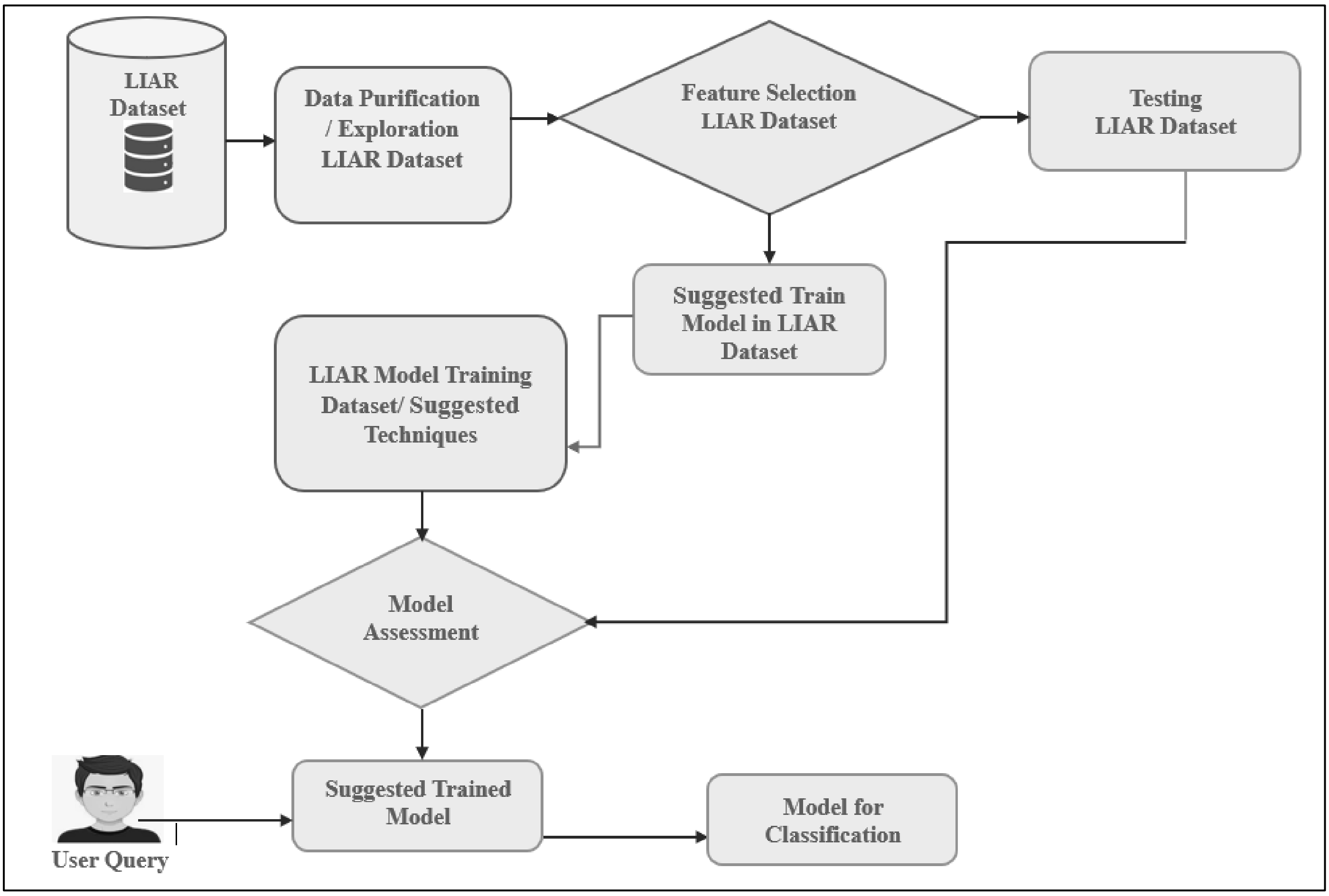

Fig 4 The Architecture of Fake News Detection Model

Hyphenation has several purposes in English, such as indicating word grouping after noisy vowel pairings and connecting nouns as names. While the first example may be easily understood as a single token (commonly written as coeducation), the second one has to be broken down into many words, leaving the intermediate position unclear. Text cleaning procedures make use of NLTK (Natural Language Toolkit) technology. Making a list of words from raw text is made easier with its help [83]. It learns words by dividing texts into their component parts and selecting strings of alphanumeric characters (a-z, A-Z, 0-9, and '_'). Following this, it gets rid of all whitespace, quote marks, and commas. The frequency of a term inside a text, specific to each word and document, is known as term frequency (TF). One way to measure a document's importance inside a collection is by looking at its document frequency (DF). Factor frequency (TF) measures how often a word appears in d papers, while document frequency (DF) counts how many times a phrase appears in N documents [85]. To help determine the informativeness of term t, we have IDF, which is the opposite of DF. For frequently occurring stop words in the raw text, the IDF value will be modest [86].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

IX. OUTCOME EVALUATION

Here, we'll take a look at the data via an analysis of the dataset, and then we'll talk about how well this technique worked. For each class, the initial statistical analysis is to count the number of words per sentence. False and real CSV files are both included in the dataset [70]. There are 44,898 rows and 4 columns in the dataset. First, as shown in Table 2, we calculate the TFIDF score for every word in the text. Plus, each sentence is treated as its own text item, and the content is further divided into phrases [87]. The combined POS-tagged text and the sanitized text both undergo these processes. At this point, we use classification algorithms to determine if news stories are fake or real [88].

Table 2 The Model's Performance of Classification Outcomes for the Test LIAR Datasets

Machine Learning Algorithms	Performance Summary for 80% - 20%	
	Imperfection Rate	Precision
Logistic Regression	2.39%	97.61 %
Long Short-Term Memory (LSTM)	4.70%	95.30%
Random Forest	3.89%	96.11 %
Decision Tree (DT)	7.82%	92.18%
Convolutional Neural Network (CNN)	2.90%	97.10%
Recurrent Neural Network (RNN)	7.30%	92.70%
Ada Boost	11.55%	88.45%

The algorithms' total precise accuracy is shown in Figure 5. A percentage is used to represent the algorithm's accuracy. The results show that the algorithm has a good success rate. The error rate, or the frequency of misclassifications, is a measure of the algorithm's performance. To determine it, you may use this equation.

Imperfection Rate = 100 – Precision

The defined attribute will be the output-specific attribute, while the term frequencies and count vectorizers are obtained and utilized as input-specific features for the given classification model. The TF-IDF approach, vectorizer count, and classification are all integrated via a pipeline manner in this methodology. A data sequence is transformed and linked inside a single model before being tested to get results. The training time is reduced since this approach executes many processes simultaneously, such as scraping and classification.

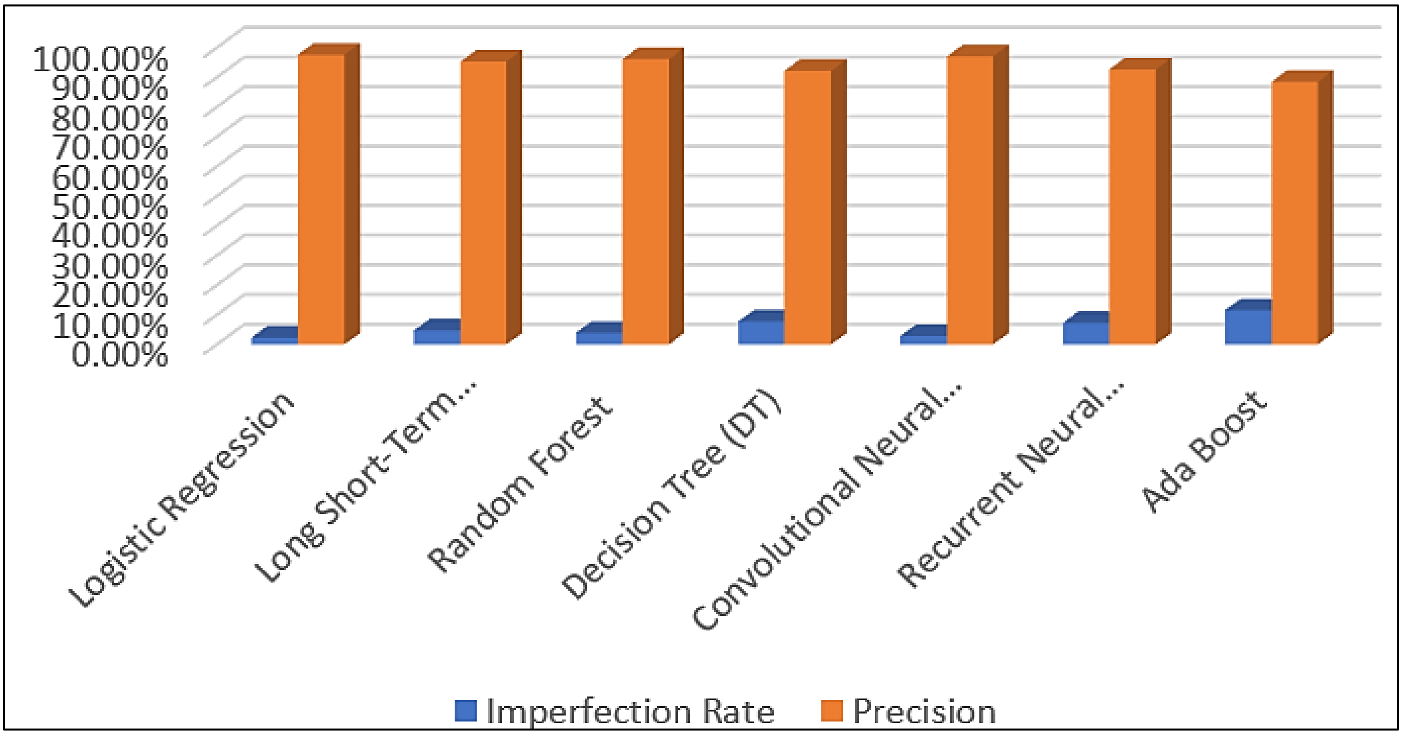

Fig 5 The Model's Performance Summary LIAR Datasets

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

➢ *Performance Indicators*

To detect bogus news, the proposed technique was provided with a substantial quantity of news items [90]. Upon obtaining the classification results using the confusion matrix, the performance metrics accuracy (A), precision (P), recall (R), and F-measure (F) were assessed using the following equations.

$$\textbf{Accuracy (A)} = \frac{T_p + T_n}{T_p + F_p + T_n + F_n}$$

$$\textbf{Precision (P)} = \frac{T_p}{T_p + F_p}$$

$$\textbf{Recall (R)} = \frac{T_p}{T_p + F_n}$$

$$\textbf{F-Measure (F)} = \frac{2 * P * R}{P + R}$$

In this section, Tp denotes the total count of news items accurately classified as favorable for a certain news [91] category. The negative value shown in figure 6 signifies the amount of news items accurately classified as detrimental for a certain news category, Tn.

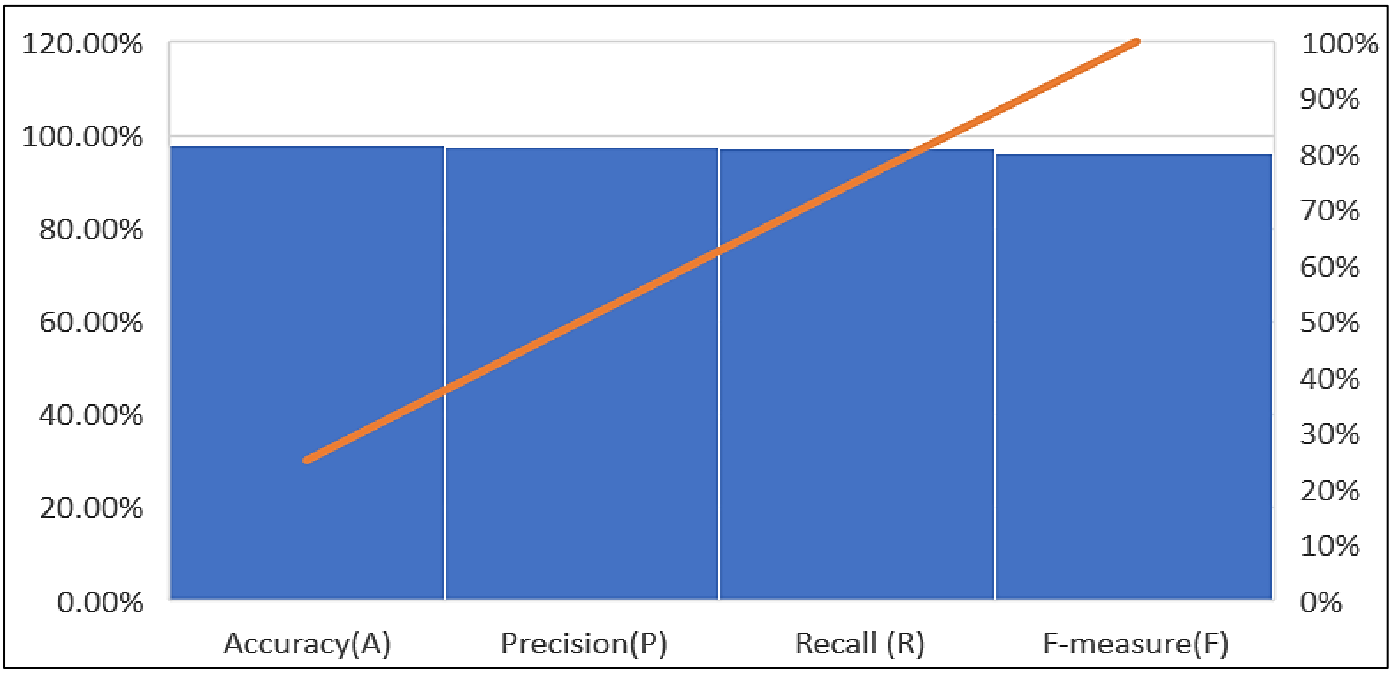

Fig 6 The Proposed Model Performance Metrics Accuracy(A), Precision(P), Recall(R), and F-measure(F)

"False positive value" (Fp) refers to the number of news items that are incorrectly identified but nonetheless placed in the allocated category. The datasets with different samples undergo many iterations [92]. As shown in Table 1, the results are then averaged and compared using the evaluation criteria that were previously mentioned [93]. According to the results of the evaluation, this method is far better than the alternatives [94].

X. CONCLUSION

The news landscape in the modern digital era has shifted from print media to social media platforms. The fast and unfiltered distribution of news on social media platforms has led to the tremendous proliferation of false information. The spread of disinformation poses a serious threat on a worldwide scale. Hence, it is crucial for the social, political, and economic spheres to be adequately equipped to comprehend and identify the vast, pressing, and diverse disinformation that is spread every day. LIAR's massive size makes it possible to refine statistical and computational approaches to identifying false news. Using LIAR's genuine, real-world, brief statements from different speakers and circumstances, researchers are able to build a more complete false news detector. According to the study, the fake news classifier had an accuracy rate of 97.61%. Results have been substantial and gratifying using the proposed method. If future research wants to improve the accuracy of fake news classification beyond 97.61%, additional discriminators should be utilized. There is room for improvement, even if the proposed method performs better than the alternatives. According to the findings of the experiment, the suggested model outperformed all other classifiers when it came to predicting the spread of fake news. For every approach, we have tested the classifier models' recall, accuracy, precision, and F-measures.

➢ *Data Availability*

The study used open-source dataset and is accessed from the weblink https://paperswithcode.com/dataset/liar

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

REFERENCES

[1]. Zhou, X., Reza, Z., Kai S., Huan, L. (2019). Fake news: Fundamental theories, detection strategies and challenges. Twelfth ACM International Conference on Web Search and Data Mining, pp. 836-837

[2]. Ms Farah Shan, Versha Verma, Apoorva Dwivedi, Y. Perwej, Ashish Kumar Srivastava, "Novel Approaches to Detect Phony Profile on Online Social Networks (OSNs) Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 9, Issue 3, Pages 555-568, May-June 2023-2023, DOI: 10.32628/CSEIT23903126

[3]. Dong, X., Victor U., Qian, L. (2020). Two-path deep semi supervised learning for timely fake news detection. IEEE Transactions on Computational Social Sys., 7(6): 1386 1398

[4]. Xu, K., Wang, F., Wang, H., Yang, B. (2020). Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Science and Technology, 25(1): 20-27

[5]. De Beer, D.; Matthee, M. Approaches to identify fake news: A systematic literature review. In International Conference on Integrated Science, Cambodia; Springer: Basel, Switzerland, pp. 13–22, 2020

[6]. Sachin Bhardwaj, Apoorva Dwivedi, Ashutosh Pandey, Y. Perwej, Pervez Rauf Khan, "Machine Learning-Based Crowd Behavior Analysis and Forecasting", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 9, Issue 3, Pages 418-429, May-June 2023-2023, DOI: 10.32628/CSEIT23903104

[7]. Chiang, T.H.C.; Liao, C.-S.; Wang, W.-C. Investigating the Difference of Fake News Source Credibility Recognition between ANN and BERT Algorithms in Artificial Intelligence. Appl. Sci., 12, 7725, 2022

[8]. Goldani, M.H.; Momtazi, S.; Safabakhsh, R. Detecting fake news with capsule neural networks. Appl. Soft Comput. 101, 106991, 2021

[9]. Bühler, J.; Murawski, M.; Darvish, M.; Bick, M. Developing a Model to Measure Fake News Detection Literacy of Social Media Users. In Disinformation, Misinformation, and Fake News in Social Media; Springer: Basel, Switzerland, pp. 213–227, 2020

[10]. Apoorva Dwivedi, Dr. Basant Ballabh Dumka, Susheel Kumar, Dr. Fokrul Alom Mazarbhuiya, Ms Farah Shan, Y. Perwej, "State of the Art Machine Learning Techniques for Detecting Fake News", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN: 2395-1990, Online ISSN: 2394-4099, Volume 10, Issue 4, Pages 115-130, July-August 2023, DOI: 10.32628/IJSRSET23103191

[11]. Sachin Bhardwaj, Apoorva Dwivedi, Ashutosh Pandey, Y. Perwej, Pervez Rauf Khan, "Machine Learning-Based Crowd Behavior Analysis and Forecasting", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 9, Issue 3, Pages 418-429, May-June 2023-2023, DOI: 10.32628/CSEIT23903104

[12]. Farghaly, A.; Shaalan, K.: Arabic natural language processing: challenges and solutions. ACM Trans. Asian Lang. Inf. Process. 8(4), 1–22, 2009

[13]. A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE International Conference on Com. Vi., pages 1–11, 1, 3, 2019

[14]. Á. Figueira and L. Oliveira, "The current state of fake news: Challenges and opportunities," Procedia Computer Science, vol. 121, pp. 817–825, 2017

[15]. Jiang T, Li JP, Haq AU, Saboor A, Ali A,"A novel stacking approach for accurate detection of fake news", IEEE Access 9:22626–22639, 2021

[16]. Y. Perwej, "An Optimal Approach to Edge Detection Using Fuzzy Rule and Sobel Method", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), ISSN (Print) : 2320 – 3765, ISSN (Online): 2278 – 8875, Volume 4, Issue 11, Pages 9161-9179, 2015, DOI: 10.15662/IJAREEIE.2015.0411054

[17]. Chen W, Zhang Y, Yeo CK, Lau CT, Sung Lee B,"Unsupervised rumor detection based on users' behaviors using neural networks", Pattern Rec. Lett 105:226–233, 2018

[18]. Y. Perwej, "Recurrent Neural Network Method in Arabic Words Recognition System", International Journal of Computer Science and Telecommunications (IJCST), Sysbase Solution (Ltd), UK, London, (http://www.ijcst.org) , ISSN 2047-3338, Volume 3, Issue 11, Pages 43-48, 2012

[19]. Farghaly, A.; Shaalan, K.: Arabic natural language processing: challenges and solutions. ACM Trans. Asian Lang. Inf. Process. 8(4), 1–22, 2009

[20]. Bhavesh Kumar Jaisawal, Y. Perwej, Sanjay Kumar Singh, Susheel Kumar, Jai Pratap Dixit, Niraj Kumar Singh, "An Empirical Investigation of Human Identity Verification Methods" , International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Volume 10, Issue 1, Pages 16-38, 2022, DOI: 10.32628/IJSRSET2310012

[21]. Xu, K., Wang, F., Wang, H., Yang, B. (2020). Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Science and Technology, 25(1): 20-27

[22]. Al Zaatari, Ayman and El Ballouli, Rim and ELbassouni, Shady and El-Hajj, Wassim and Hajj, Hazem and Shaban, Khaled and Habash, Nizar and Yahya, Emad. (2016) "Arabic corpora for credibility analysis." Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) 4396–4401

[23]. Asif Perwej, Y. Perwej, Nikhat Akhtar, and Firoj Parwej, "A FLANN and RBF with PSO Viewpoint to Identify a Model for Competent Forecasting Bombay Stock Exchange", COMPUSOFT, SCOPUS, An

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

International Journal of Advanced Computer Technology, 4 (1), Volume-IV, Issue-I, Pages 1454-1461, 2015, DOI: 10.6084/ijact.v4i1.60

[24]. Allcott H, Gentzkow M," Social media and fake news in the 2016 election", J Econ Perspect 31(2):211–36, 2017

[25]. Jin Z, Cao J, Zhang Y, Zhou J, Tian Q." Novel visual and statistical image features for microblogs news verification", IEEE Trans Multimed 19(3):598–608, 2016

[26]. Y. Perwej, Firoj Parwej, Asif Perwej, "Copyright Protection of Digital Images Using Robust Watermarking Based on Joint DLT and DWT", International Journal of Scientific & Engineering Research (IJSER), France, ISSN 2229-5518, Volume 3, Issue 6, Pages 1- 9, 2012

[27]. Y. Perwej, Asif Perwej, Firoj Parwej, "An Adaptive Watermarking Technique for the copyright of digital images and Digital Image Protection", International journal of Multimedia & Its Applications (IJMA), Academy & Industry Research Collaboration Center (AIRCC), USA, Volume 4, No.2, Pages 21- 38, 2012, DOI: 10.5121/ijma.2012.4202

[28]. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19, 1 (2017), 22–36

[29]. Elhadad MK, Li KF," Detecting misleading information on COVID-19", IEEE Access 8:165201–165215, 2020

[30]. Potthast, Martin and Kopsel, Sebastian and Stein, Benno and Hagen, Matthias., "Clickbait detection." ¨ European Conference on Information Retrieval 810–817, 2016

[31]. Nikhat Akhtar, Devendera Agarwal, "An Efficient Mining for Recommendation System for Academics", International Journal of Recent Technology and Engineering (IJRTE), ISSN 2277-3878 (online), SCOPUS, Volume-8, Issue-5, Pages 1619-1626, 2020 , DOI: 10.35940/ijrte.E5924.018520

[32]. Y. Perwej, "Unsupervised Feature Learning for Text Pattern Analysis with Emotional Data Collection: A Novel System for Big Data Analytics", IEEE International Conference on Advanced computing Technologies & Applications (ICACTA'22), SCOPUS, IEEE No: #54488 ISBN No Xplore: 978-1-6654-9515-8, Coimbatore, India, 4-5 March 2022, DOI: 10.1109/ICACTA54488.2022.9753501

[33]. Zhou, X., Zafarani, R.: A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput. Surv. (CSUR) 53(5), 1–40. 2020

[34]. Neves JC et al,"GANprintR: improved fakes and evaluation of the state of the art in face manipulation detection", IEEE J Sel Top Signal Proc 14(5):1038–1048, 2020

[35]. Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey. ACM Computing Surveys (CSUR) 51, 2, 32, 2018

[36]. Y. Perwej, Shaikh Abdul Hannan, Nikhat Akhtar, "The State-of-the-Art Handwritten Recognition of Arabic Script Using Simplified Fuzzy ARTMAP and Hidden Markov Models", International Journal of Computer Science and Telecommunications (IJCST), Sysbase Solution (Ltd), UK, London, ISSN 2047-3338, Volume, Issue 8, Pages 26 - 32, 2014

[37]. Zhou X, Zafarani R ,"Fake news: a survey of research, detection methods, and opportunities", 2018,

[38]. Wang Y, Ma F, Jin Z, Yuan Y, Xun G, Jha K, Su L, Gao J (2018) Eann: Event adversarial neural networks for multi-modal fake news detection. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 849–857

[39]. Jin Z, Cao J, Guo H, Zhang Y, Luo J, Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM international conference on Multimedia, pp 795–816, 2017

[40]. Wang WY Liar, liar pants on fire: A new benchmark dataset for fake news detection. In: Proceedings of the 55th annual meeting of the association for computational linguistics (vol 2: short Papers), pp 422–426, 2017

[41]. Y. Perwej, "The Bidirectional Long-Short-Term Memory Neural Network based Word Retrieval for Arabic Documents", Transactions on Machine Learning and Artificial Intelligence (TMLAI), Society for Science and Education, United Kingdom (UK), ISSN 2054-7390, Volume 3, Issue 1, Pages 16 - 27, 2015, DOI: 10.14738/tmlai.31.863

[42]. Yang F, Liu Y, Xiaohui Y, Yang M , Automatic detection of rumor on Sina Weibo. In: Proceedings of the ACM SIGKDD workshop on mining data semantics, pp 1–7, 2012

[43]. Y. Perwej, Nikhat Akhtar, Firoj Parwej, "A Technological Perspective of Blockchain Security", International Journal of Recent Scientific Research (IJRSR), ISSN: 0976-3031, Volume 9, Issue 11, (A), Pages 29472 – 29493, 2018. DOI: 10.24327/ijrsr.2018.0911.2869

[44]. A. Perwej, Prof. (Dr.) K. P. Yadav, Prof. (Dr.) Vishal Sood, Y. Perwej, " An Evolutionary Approach to Bombay Stock Exchange Prediction with Deep Learning Technique", IOSR Journal of Business and Management (IOSR-JBM), e-ISSN: 2278-487X, p-ISSN: 2319-7668, USA, Volume 20, Issue 12, Ver. V, Pages 63-79, 2018, DOI: 10.9790/487X-2012056379

[45]. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explor. Newslett., 19, 22–36, 2017

[46]. [2] E. C. T. Jr., Z. W. Lim, and R. Ling, "Defining "fake news"," Digital Journalism, vol. 6, no. 2, pp. 137–153, 2018. doi: 10.1080/21670811.2017.1360143

[47]. Y. Perwej, Kashiful Haq, Uruj Jaleel, Firoj Perwej, "Block Ciphering in KSA, A Major Breakthrough in Cryptography Analysis in Wireless Networks", International Transactions in Mathematical Sciences

and Computer, India, ISSN-0974-5068, Volume 2, No. 2, Pages 369-385, July-December 2009

[48]. Thorne, J., Chen, M., Myrianthous, G., Pu, J., Wang, X., Vlachos, A. (2017). Fake news stance detection using stacked ensemble of classifiers. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing Meets Journalism, pp. 80-83

[49]. Shu K, Wang S, Liu H,"Exploiting tri-relationship for fake news detection", Association for the Advancement of Artifcial Intelligence, arXiv preprint arXiv:1712.07709, 2017

[50]. Jin Z, Cao J, Zhang Y, Zhou J, Tian Q," Novel visual and statistical image features for microblogs news verification", IEEE Trans Multimed 19(3):598–608, 2016

[51]. Saurabh Sahu, Km Divya, Neeta Rastogi, Puneet Kumar Yadav, Yusuf Perwej, "Sentimental Analysis on Web Scraping Using Machine Learning Method" , Journal of Information and Computational Science (JOICS), ISSN: 1548-7741, Volume 12, Issue 8, Pages 24-29, August 2022**,** DOI: 10.12733/JICS.2022/V12I08.535569.67004

[52]. Zhou X, Zafarani R, Shu K, Liu H (2019) Fake news: fundamental theories, detection strategies and challenges, In: Proceedings of the twelfth ACM international conference on web search and data mining, WSDM'19. Association for Computing Machinery, New York, NY, USA, pp 836–837

[53]. Dawar Husain, Y. Perwej, Satendra Kumar Vishwakarma, Prof. (Dr.) Shishir Rastogi, Vaishali Singh, Nikhat Akhtar, "Implementation and Statistical Analysis of De-noising Techniques for Standard Image", International Journal of Multidisciplinary Education Research (IJMER), ISSN:2277-7881, Volume 11, Issue10 (4), Pages 69-78, 2022, DOI: 10.IJMER/2022/11.10.72

[54]. Firoj Parwej, N. Akhtar, Y. Perwej, "An Empirical Analysis of Web of Things (WoT)", International Journal of Advanced Research in Computer Science (IJARCS), ISSN: 0976-5697, Volume 10, No. 3, Pages 32-40, May 2019, DOI: 10.26483/ijarcs.v10i3.6434

[55]. H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election", The Journal of Economic Perspectives, vol. 31, no. 2, pp. 211-235, 2017

[56]. Prof. Kameswara Rao Poranki, Y. Perwej, Dr. Asif Perwej," The Level of Customer Satisfaction related to GSM in India ", TIJ's Research Journal of Science & IT Management – RJSITM, International Journal's-Research Journal of Science & IT Management of Singapore, ISSN: 2251-1563, Singapore, in www.theinternationaljournal.org as RJSSM, Volume 04, Number: 03, Pages 29-36 , 2015

[57]. Guess A, Nagler J, Tucker J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances. 2019;5: eaau4586. pmid:30662946

[58]. N. Akhtar, Nazia Tabassum, Asif Perwej. Y. Perwej," Data Analytics and Visualization Using Tableau Utilitarian for COVID-19 (Coronavirus)", Global Journal of Engineering and Technology Advances (GJETA), Volume 3, Issue 2, Pages 28-50, 2020, DOI: 10.30574/gjeta.2020.3.2.0029

[59]. Kaliyar, R.K.; Goswami, A.; EchoFakeD: Improving fake news detection in social media with an efficient deep neural network. Neu. Comput. Appl., 33, 8597–8613, 2021

[60]. Golbeck, J.; Mauriello, M.; Auxier, B.; Bhanushali, K.H.; Bonk, C.; Bouzaghrane, M.A.; Buntain, C.; Chanduka, R.; Cheakalos, P.; Everett, J.B.; et al. Fake News vs Satire: A Dataset and Analysis; WebSci '18; Association for Computing Machinery: New York, NY, USA, pp. 17–21, 2018

[61]. Y. Perwej, Dr. Shaikh Abdul Hannan, Firoj Parwej**,** N. Akhtar, **"**A Posteriori Perusal of Mobile Computing", International Journal of Computer Applications Technology and Research (IJCATR), ATS (Association of Technology and Science), India, ISSN 2319–8656 (Online), Volume 3, Issue 9, Pages 569 - 578, September 2014, DOI: 10.7753/IJCATR0309.1008

[62]. Al-Mushayt O., Haq Kashiful, Y. Perwej, "Electronic-Government in Saudi Arabia; a Positive Revolution in the Peninsula", International Transactions in Applied Sciences, India, ISSN-0974-7273, Volume 1, Number 1, Pages 87-98, July-December 2009

[63]. C.-Y. Lin, T.-Y. Li and P. Chen, "An Information Visualization System to Assist News Topics Exploration with Social Media", ACMDL, July 2016

[64]. Mykhailo Granik and Volodymyr Mesyura, "Fake news detection using naive bayes classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 900-903, 2017

[65]. M. Luo, J.T. Hancock and D.M. Markowitz, "Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues", Comm. Research, vol. 49, no. 2, pp. 171-195, 2022

[66]. Niall J Conroy, Victoria L Rubin and Yimin Chen, "Automatic deception detection: Methods for finding fake news", Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1-4, 2015

[67]. Santhanam, Laura. 2017. "New poll: 70% of Americans think civility has gotten worse since Trump took office." PBS News Hour, July 3. www.pbs.org/newshour/politics/new-poll-70-americans-think-civility-gotten-worse-since-trump-took-office

[68]. Allcott, H. and Gentzkow, M.," Social media and fake news in the 2016 election", Journal of Economic Perspectives, 31(2):211–36, 2017\

[69]. Shobhit Kumar Ravi, Shivam Chaturvedi, Dr. Neeta Rastogi, Dr. Nikhat Akhtar, Dr. Yusuf Perwej, "A Framework for Voting Behavior Prediction Using Spatial Data", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Volume 10, Issue 2, Pages 19-28, 2022, DOI: 10.55524/ijircst.2022.10.2.4

[70]. William Yang Wang, "" liar liar pants on fire": A new benchmark dataset for fake news detection", arXiv, 2017
[71]. A. Al-Sideiri, Z. B. C. Cob, and S. B. M. Drus, Machine Learning Algorithms for Diabetes Prediction: A Review Paper,‖ ACM Int. Conf. Proceeding Ser., pp. 27–32, 2019, doi: 10.1145/3388218.3388231.
[72]. Y. Perwej, Dr. Ashish Chaturvedi, "Machine Recognition of Hand-Written Characters using Neural Networks", International Journal of Computer Applications (IJCA), USA, ISSN 0975 – 8887, Volume 14, No. 2, Pages 6- 9, 2011, DOI: 10.5120/1819-2380
[73]. Dr. E. Baraneetharan, Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey, ‖ J. Inf. Technol. Digit. World, vol. 02, no. 03, pp. 161–173, 2020, doi: 10.36548/jitdw.2020.3.004.
[74]. Y. Perwej, Firoj Parwej, Nikhat Akhtar, "An Intelligent Cardiac Ailment Prediction Using Efficient ROCK Algorithm and K- Means & C4.5 Algorithm", European Journal of Engineering Research and Science (EJERS), Bruxelles, Belgium, ISSN: 2506-8016 (Online), Vol. 3, No. 12, Pages 126 – 134, 2018, DOI: 10.24018/ejers.2018.3.12.989
[75]. C. Zhenhai, Liu. Wei, "Logistic Regression Model and Its Application*,"* Journal of Yanbian University (Natural Science Edition), vol. 38 (01), pp 28–32, 2012
[76]. A. Telikani, A. Tahmassebi, W. Banzhaf, and A. H. Gandomi, Evolutionary Machine Learning: A Survey, ACM Comput. Surv., vol. 54, no. 8, 2022
[77]. Wei Xiong, Bo Du, Lefei Zhang, Ruimin Hu and Dacheng Tao, "Regularizing Deep Convolutional Neural Networks with a Structured Decorrelation Constraint", IEEE 16th International Conference on Data Mining (ICDM), pp. 3366-3370, 2016
[78]. Y. Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN (Online): 2278-1021, ISSN (Print): 2319-5940, Volume 4, Issue 2, Pages 10 - 16, 2015, DOI: 10.17148/IJARCCE.2015.4203
[79]. N. Kwak, Introduction to Convolutional Neural Networks (CNNs), 2016
[80]. Z Li, C Ding, S Wang et al., "E-RNN: Design optimization for efficient recurrent neural networks in FPGAs[C]", 2019 IEEE International Symposium on High Performance Computer Architecture (HPCA), pp. 69-80, 2019
[81]. Uzun, Erdinc, "A Novel Web Scraping Approach Using the Additional Information Obtained From Web Pages", IEEE Access, 8, 2020
[82]. T. Euler, "The Token Classification Framework: A multidimensional tool for understanding and classifyingcrypto tokens.," Untitled INC, 2018. http://www.untitled-inc.com/the-token-classification-framework-a-multi-dimensional-tool-for-understanding-and-classifying-crypto-tokens/, 2020
[83]. Urafsky *D*. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition I. D. Jurafsky, J. H. Martin. — Upper Saddle River, NJ : Prentice Hall, 2008. - 988 p.
[84]. Nikhat Akhtar, "Artificial Intelligence and Machine Learning in Human Resource Management for Sales research Perspective", IEEE International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Electronic ISBN:978-1-6654-7413-9, SCOPUS, ISBN:978-1-6654-7414-6, Chennai, India, 2022, DOI: 10.1109/ICSES55317.2022.9914086
[85]. Neha Kulshrestha, N. Akhtar, Y. Perwej, "Deep Learning Models for Object Recognition and Quality Surveillance", International Conference on Emerging Trends in IoT and Computing Technologies (ICEICT-2022), ISBN 978-10324-852-49, Routledge, Taylor & Francis, CRC Press, Chapter 75, pages 508-518, Goel Institute of Technology & Management, Lucknow, 2022, DOI: 10.1201/9781003350057-75
[86]. Saurabh Sahu, Km Divya, Neeta Rastogi, Puneet Kumar Yadav, Yusuf Perwej, "Sentimental Analysis on Web Scraping Using Machine Learning Method" , Journal of Information and Computational Science (JOICS), Volume 12, Issue 8, Pages 24- 29, 2022, DOI: 10.12733/JICS.2022/V12I08.535569.67004
[87]. S. Gilda, "Notice of violation of ieee publication principles: Evaluating machine learning algorithms for fake news detection", 2017 IEEE 15 th student conference on research and development, IEEE , pp. 110-115, 2017
[88]. Y. Perwej, "An Optimal Approach to Edge Detection Using Fuzzy Rule and Sobel Method", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), ISSN (Print) : 2320 – 3765, ISSN (Online): 2278 – 8875, Volume 4, Issue 11, Pages 9161-9179, 2015, DOI: 10.15662/IJAREEIE.2015.0411054
[89]. Y. Perwej, Dr. Ashish Chaturvedi, "Machine Recognition of Hand Written Characters using Neural Networks", International Journal of Computer Applications (IJCA), USA, ISSN 0975 – 8887, Volume 14, No. 2, Pages 6- 9, 2011, DOI: 10.5120/1819-2380
[90]. Mr. Vinay Kumar, Neha Goyal, Yusuf Perwej, Devendra Agarwal, Alok Mishra, Prachi Chauhan," Text Based Data Extraction (TBDE) Approach from Different Data-Sets Source", Emerging Trends in IoT and Computing Technologies, 1st Edition, eBook ISBN No. 978-1-032-87924-6, SCOPUS, CRC Press, Taylor & Francis, London, Pages 196- 202, Published 2024 Link: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003535423-34/text-based-dataextraction-tbde-approach-different-data-sets-source-vinay-kumar-neha-goyal-yusuf-perwej-devendraagarwal-alok-mishra-prachi-chauhan?context=ubx&refId=06b2c833-9001-4024-

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1003

9e26-e4954d76a5be DOI: 10.1201/9781003535423-33

[91]. Nagoudi, E.M.B.; Elmadany, A.R.; Abdul-Mageed, M.; Alhindi, T.; Cavusoglu, H.: Machine generation and detection of Arabic manipulated and fake news. arXiv, pp. 69–84, 2020

[92]. N. Akhtar, Devendera Agarwal, "An Efficient Mining for Recommendation System for Academics", International Journal of Recent Technology and Engineering (IJRTE), ISSN 2277-3878 (online), SCOPUS, Volume-8, Issue-5, Pages 1619-1626, 2020, DOI: 10.35940/ijrte.E5924.018520

[93]. Elhadad, M.K.; Li, K.F.; Gebali, F. Detecting misleading information on COVI.-19.IEEE Access, 165201,165215, 2020

[94]. Knshnan, S.; Chen, M.: Identifying tweets with fake news. In: Proceedings 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science. IRI 2018, vol. 67, pp. 460–464, 2018