Fake News Detection on Kashmir Issue Using Machine Learning Techniques

Misbah Kazmi[1], Sadia Nauman[2], Sadaf Abdul Rauf[3],
S. Ali[4], Ali Daud[5], Bader Alshemaimri[6]
Department of Computer Science, Fatima Jinnah Women University, Rawalpindi, Pakistan[1,3]
Department of Communication and Media Studies, Fatima Jinnah Women University, Rawalpindi, Pakistan[2]
Department of Computer Science, National University of Science and Technology, Islamabad, Pakistan[4]
Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates[5]
Software Engineering Department, King Saud University, Riyadh, Saudi Arabia[6]

***Abstract*—Focusing events are sudden, impactful occurrences that spark widespread discussions. Analyzing fake news during such events is challenging due to limited and short-lived datasets. Online fact checkers are slow in identifying fake news, and internet communities and forums become the primary source of news, allowing unchecked dissemination. This study proposes a machine learning approach to predict fake news during the revocation of Article 370 in Kashmir as a focusing event. Small dataset from 20th August till 2nd September is collected and user profile parameters are utilized for effective classification. Five classifiers were employed, with Random-Forest and Logistic-Regression achieving the highest F1 scores of 74 per cent. Results identifies prevalent words in true and false news tweets, aiding in fake news detection. This approach mitigates misinformation during events with limited data, contributing to a reliable online environment. The research is valuable for major geopolitical shifts, natural disasters, and social movements.**

***Keywords*—Classification algorithm; fake news; Kashmir issue; machine learning techniques**

I. Introduction

Focusing events are short lived events that capture public attention and direct it towards a specific aspect of a problem [1]. They can sway politics by mobilizing public support for one side of a debate, potentially influencing political outcomes [6]. Internet communities and forums have become a powerful and cost-effective platform for sharing news, leading to the dissemination of both real and fake information [32], [11]. Fake news detection involves predicting misleading articles intended to deceive readers [40], [62]. During Hurricane Sandy, a study found that 86 per cent of tweets sharing fake images were re-tweets, illustrating the alarming speed at which misinformation spreads [44].

Fake news aims to mislead readers by presenting false information, posing a high risk as it is driven by motives such as social popularity, political agendas, or manipulation of public opinion [43] [39] [50]. Detecting fake news on social media is an emerging research field of great importance to mitigate its adverse impact on the public and the news ecosystem [42].

However, existing research on fake news detection [59], [60], [26] has primarily focused on general-purpose datasets or misinformation in stable, long-term contexts (e.g., elections or health crises). Very limited studies explore detection during focusing events — short, high-impact political developments where data is sparse and rapidly evolving. This creates a significant gap in developing real-time fake news detection methods tailored for such scenarios [19]. This leaves an important gap in the development of real-time fake news detection techniques that are specific to these situations. Our research study tries to address this gap by applying conventional machine learning methods to short-term time series data on Twitter, focusing on the revocation of Article 370 in Kashmir.

In the field of natural language processing (NLP) and computational linguistics, Machine learning(ML) and Deep learning(DL) techniques are being used as powerful analysis tools for identification, classification and detection of language related issues [51] [41]. Our research focuses on Article 370 revocation by India in the state of Jammu and Kashmir as a significant event [29]. In August 2019, the Indian government abolished Article 370, bringing the region under direct central government control [8]. The event generated extensive discussion on Twitter, with a bombardment of visual and textual data [18]. For scenarios like focusing events, where limited data size is available. Traditional ML techniques are effective in classification problems, as DL models require a large amount of training data to produce quality results [40], [59]. Recently, these techniques have been explored as the starting point towards the automation of task selection mechanism of Crowd-Sourced software development (CSSD) platforms [45].

The research question in focus in this study is: What are the performance levels of traditional machine learning models in identifying fake news in high-impact short-duration political events when applied to event-specific social media data?

This study utilizes User profiles on the Twitter social network, for measuring fake news identification using explicit and implicit features [41] [53]. For news tweet dataset, data of twitter users responses and news sources on the revocation of Article 370 in Kashmir was collected. Motivation behind the data collection is identification of fake and real news spread across the platform in the backdrop of the event. Data collection period ranged from 20th August to 02 September, 2019 [52]. The credibility of collected tweet news was checked using Google News search engine, through valid sources and parameters, such as the presence of a "Verified" symbol on twitter account sharing the news [47], friends and following counts of the twitter account [16], re-tweet counts, tweet

Algorithm 1 Fake News Tweet Detection using Logistic Regression

```
for Eachtweet = 1, 2, ... do
    for tCount ← ∑(numberofwords) do
        Tokenize tweet by word
    end for
    Apply pre-processing techniques
    Tweet ⇐ words − (stopword U punctuationword)
    Tweet = Tweet.wordNet Lemmatizer(words)
    for EachTweet do
        Tweet classify.logisticRegression(TweetHyperParams)
        if Tweet = True then
            TweetLabel ⇐ real
        else if Tweet = False then
            TweetLabel ⇐ fake
        end if
    end for
end for
```

creation date, and user reviews about the given tweet [10] [36] [13].

This study introduces a system that automatically detects and evaluates fake news on Twitter, specifically focusing on the revocation of Article 370 in Kashmir. The system captures tweets from various sources and utilizes profile parameters to classify them as fake or real. It trains a model by extracting relevant features and employs selected classifiers. F1 score is used to evaluate model performance.

Our research primarily contributed in:

- Affirming the credibility of traditional ML techniques to analyse “focusing events”. The framework identifies fake tweets from different sources and enhance the ability to discern the credibility of news.
- Classification of the most prevalent words in true and false news tweets, highlighting the linguistic patterns that can aid in detecting fake news.

Our study integrates various ML techniques to effectively detect and evaluate fake news on Twitter. We compared five different ML algorithms based on their F1 scores and accuracy. The study presents research outcomes, system limitations, and utilizes word clouds and statistics to classify prevalent words in news tweets, highlighting linguistic patterns for fake news detection, the process is shown in Algorithm 1 on the task of fake news detection.

The rest of the study is organized as follows. The background information about the Kashmir conflict, fake news, and machine learning methods are provided in Section II. The proposed methodology such as data collection, feature extraction and classification algorithms are explained Section III. The results are discussed in Section IV and compared in terms of classifier performance. The study ends with the limitations and future directions in Section V.

II. Background

The related work can be divided into three sub-sections; Conflict of Jammu and Kashmir and revocation of Article 370 of Indian constitution as its focusing event, Fake news and its widespread impact in the backdrop of such events and lastly ML techniques as effective approaches for mitigating the spread and impact of fake news.

A. Conflict of Jammu and Kashmir

Territory of Jammu and Kashmir has been a source of contention between Pakistan and India for over 70 years, leading to four wars in 1948, 1965, 1971, and 1999[56]. Additionally, various groups within Kashmir have their own agenda, which span from seeking affirmative discrimination to advocating for a separate political status or even outright secession [38]. This intricate mix of factors contributes to the intricate nature of the Kashmir conflict.

In August 2019, the Indian government made a significant move by revoking Article 370 of the Indian Constitution. This decision eliminated region’s separate constitution and flag, bringing it under the direct control of the central government of India. The decision received both support and opposition, sparking debates about its impact on the region’s identity, security, and relations with the rest of India [8]. The event triggered a widespread response across the world and on social media, particularly on Twitter, with an influx of visual and textual data [18].

B. Fake News and its Impact

Cognitive theories [15], [20] suggest that due to our inherent gullibility, people are prone to believing information that aligns with their pre-existing views (confirmation bias). Consequently, individuals may readily share such information without proper verification, further distorting facts that do not align with their beliefs. In the past, foreign actors primarily used fake news for political manipulation. However, it is now increasingly being leveraged across various domains [2] [25] to influence public opinion. During COVID-19 pandemic, misleading and fake information spread all across social media which resulted in false claims and a surge in pseudo scientific health treatments related to the virus and potential remedies [7]. The landscape of fake news has undergone a significant transformation in recent years[55], [35]. Traditionally, the dissemination of misinformation was primarily through conventional media channels. However, the rise of online news platforms and social media has drastically altered the ecosystem of fake news [4] [33] [27]. Hence, the detection of fake news has garnered significant attention in recent years.

C. Machine Learning Techniques

Previously, ML models are being used for prediction problems [30] [37] to eventually help organizations with decision making and identifying future trends. [10] proposed a system utilizing 45 features to predict news accuracy, capturing tweet author specifications, tweet stream properties, contextual aspects, and trends. In [22], the authors proposed a modest approach to fake news detection using the naive Bayes classifier, incorporating user accounts, reverse image searching, and cross-verification of sources. In [12], the authors discussed DL algorithms such as RNN and LSTMs. In Twitter account and bot detection, [14] introduced a system using Support Vector Machine Neural Network (SVM-NN) and other ML

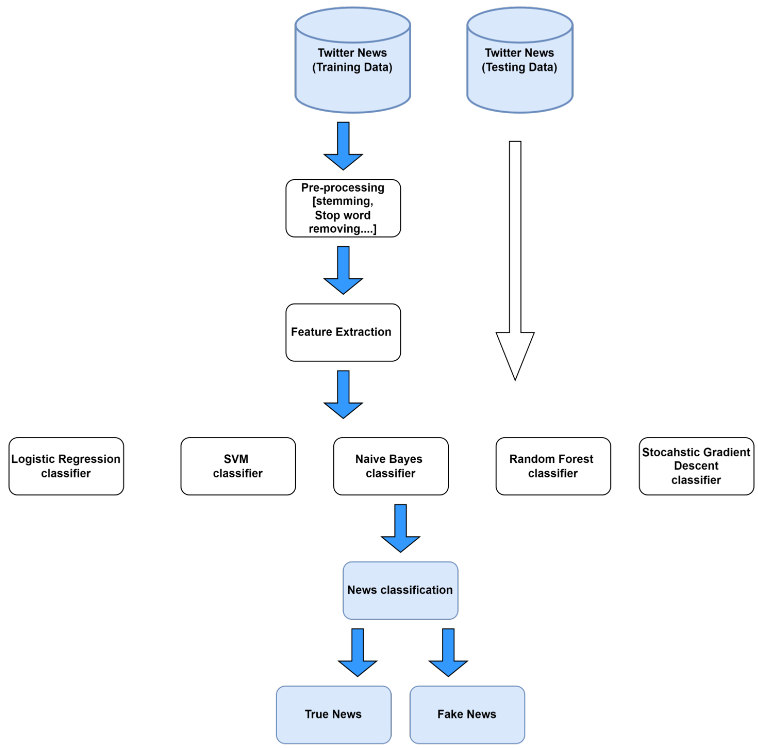

Fig. 1. Machine learning based process of fake news classification.

algorithms. In [17], [46], and [21], the authors explored similar techniques for detecting fake accounts. In [41], the authors measured trust levels of users to identify fake news, employing explicit profile features. In [49], the authors proposes an improved decision tree by changing threshold for their bi-criteria optimization. These improved decision trees are useful for fake news detection.Recently, [45] [54] proposes multi-modal fake news detection using ensemble learning. Modal exploits publicly available dataset and uses NLP techniques for textual information pre-proccessing and gauging the sentiment from each news.

III. Proposed System

An initial set of 6570 general tweets on the revocation of article 370 was collected. During the pre-processing step, tweet redundancy was removed and relevant tweet news content was kept. Graphical data was excluded from dataset due to the challenges in determining the authenticity of images and the complexity involved in assessing hidden objects within them [51], [24]. As a result, 402 news tweets are obtained for model training. Tweet news dataset was filtered to include English news content and remove any unwanted columns and empty rows. Before training, some hyper parameters are fixed for classification and filtered dataset was distributed into train and test data. Table I provides initial distribution of our real and fake news tweets dataset. Fig. 1 shows a high-level overview of the machine learning-based process used to classify fake news.

A. Data Pre-Processing

Fig. 1 illustrates the classification process of the proposed system. After tweet news dataset collection, data pre-processing techniques are applied to remove URLs, extra spaces, hashtags, and other intricacies. The tweets are further refined using stop word and punctuation removal, sentence segmentation, lower casing, and tokenization. These techniques removed all data redundancies and transformed tweets into a unified form for ML algorithms [23]. Additionally, N-gram features were extracted and using TF (Term Frequency) and TF-IDF (Term Frequency-Inverted Document Frequency) a feature matrix is formed. At the end, results of five different ML algorithms - Stochastic Gradient Descent (SGD), Logistic Regression (LR), Random Forest (RF) classifiers, Linear SVM and Naive-Bayes (NB) were compared based on F1 score and accuracy using the Python Natural Language Toolkit (NLTK).

Main steps of data pre-processing techniques being used are:

1) Splitting: This step involves breaking a text document into individual words or tokens [9].

2) Stop words removing: Stop words are used words with no significant meaning in given context. Removing stop words reduces noise and improve the efficiency of text analysis [17].

3) Stemming: Stemming reduce the words to the word base form called “Stem”. For classification method speed and accuracy, we used the commonly used Porter stemmer [28].

4) N-gram model: N-gram model represents a sequence of two words or 2-grams. N-grams store spatial information and serve as an ordered document representation [10]. In proposed methodology, Bag of words and N-grams are used to represent the documents and do the classification [48]. An N-gram classifier is built and used to differentiate between fake and real news tweets.

5) Features extraction method: However, high dimensionality is a challenge for text categorization since documents which often contain large number of words and phrases. To address this issue, two primary feature selection methods are investigated: Term Frequency (TF) and Term Frequency-Inverted Document Frequency (TF-IDF) [31]. They can mitigate the impact of redundant and irrelevant features on classifier accuracy and performance [34].

B. Prediction Algorithm

Prediction model for fake tweet news identification is built by implementation of five ML classifiers from scratch. These algorithms include LR model, NB classifier model, SVM model, RF model, and Stochastic Gradient Descent. Classification and prediction of fake news was checked across each classifier. To achieve consistency and reproducibility of experiments, we applied standardized hyperparameter configurations to all machine learning models, which are the default parameters in scikit-learn.

C. Tweet Analysis

To understand the distribution of real and fake news feature over the dataset, the study on each feature has been made using tools and packages such as seaborn, wordcloud and NLTK. Fig. 2 shows the total distribution of extracted tweets from 20th August to 30th August, 2019. A higher frequency trend of tweets is visible between 23rd August and 30th August as well as the frequency of true and false news tweets during this time period are visualized in Fig. 3. Table I illustrates the distribution of followers and re-tweet counts of the twitter user and created tweet respectively in true and false news.

TABLE I. FAKE AND REAL TWEETS DISTRIBUTION

Label	Count
True	175
False	125

It is evident from the chart that false news tends to have more followers compared to true news, indicating a higher level of dissemination for false information. Overall, these analysis help to characterize (and patterns) real and fake news in the included tweet news dataset. The analysis also represent a baseline for sentiment prediction of twitter users when a focusing event occurs [3].

D. Most Repeated Words in Extracted Tweets

Using word clouds, display the most frequent words related to the Kashmir conflict. Additionally, separate word clouds are created for true and false news categories in training data, as depicted in Fig. 4, respectively.

E. Tweet with more Re-tweets and Favourite Counts

Within the corpus, tweet with most number of favourite counts and re-tweets was selected and visualised using numpy max function as shown in Table II.

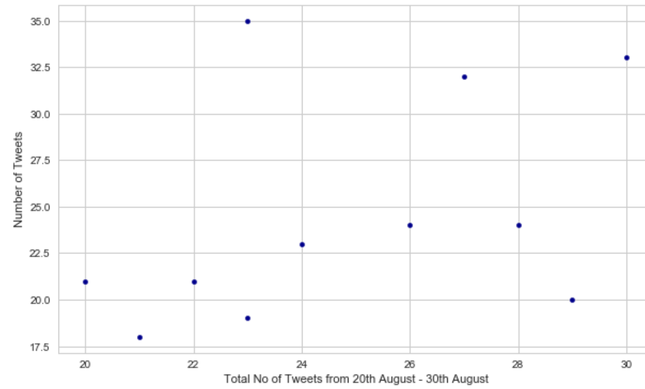

Fig. 2. Total number of tweets from 20th to 30th August.

TABLE II. TWEET HAVING MORE FAVOURITE COUNTS AND RE-TWEETS

The tweet with more likes is:
"Foreign Journals need permit to work in Kashmir. So all the hostile anti-india coverage in Western liberal media has been feed to them by their minions in India". Number of likes: 168
The tweet with more re-tweets is:
"Pakistani Envoy to U.S says world should take notice to avoid new conflict with India over Kashmir". Number of re-tweets: 488

F. Most Engagement of Twitter Users in a Given Time

Level of user engagement is identified using python 2D plotting library Matplotlib to plot a time series of tweet between given date and number of likes and re-tweets within news dataset. As per Fig. 5 , the most favourite tweet counted is between 23rd and 30th August.

G. Sources

To identify the geolocation of the tweeter users creating the tweet, different sources of extracted tweets are plotted in Fig. 6. Pakistan with the ratio of 24.52 is the The highest generator of relevant tweets with India being second highest generator of relevant tweets i.e, 23.56. Fig. 7 provides break down of true and false tweets generated from these location.

H. Results

We conducted experimentation and analysis on a dataset of 6,570 tweets relating to the revocation of Article 370 of Indian constitution by India in disputed territory of Jammu and Kashmir. After data cleaning, removing redundancy and non English tweets, five classifier models were trained on a refined dataset of 402 tweets. The overall efficiency of our classifiers is presented in Table III. We compared the F1 scores achieved by our classifiers and obtained the highest score of 74 per cent by RF classifier and LR classifier. In terms of accuracy, the RF classifier and LR classifier also achieved the highest accuracy of 82 per cent.

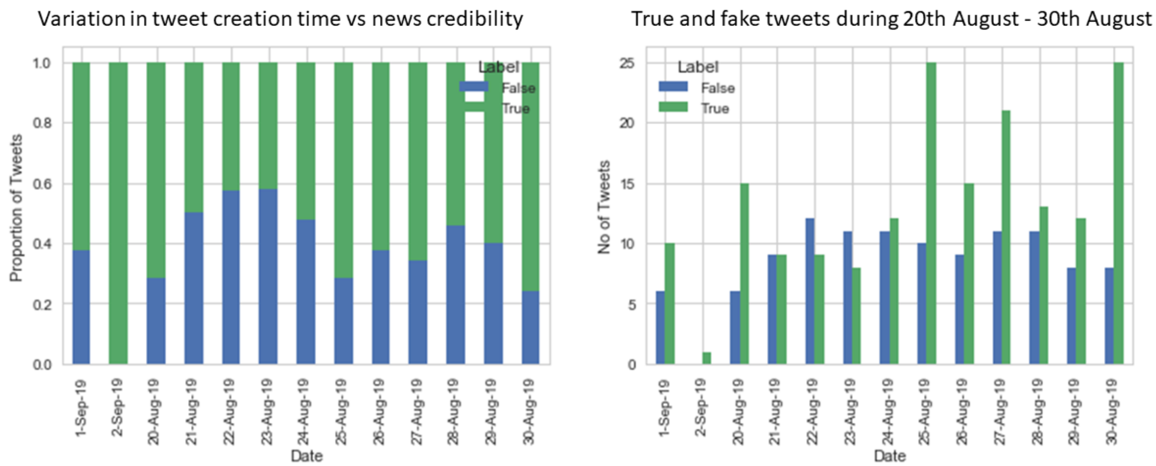

Fig. 3. Analyzing temporal variations in twitter time creation: News credibility and spread of True versus Fake tweets (Aug 20 to Aug 30).

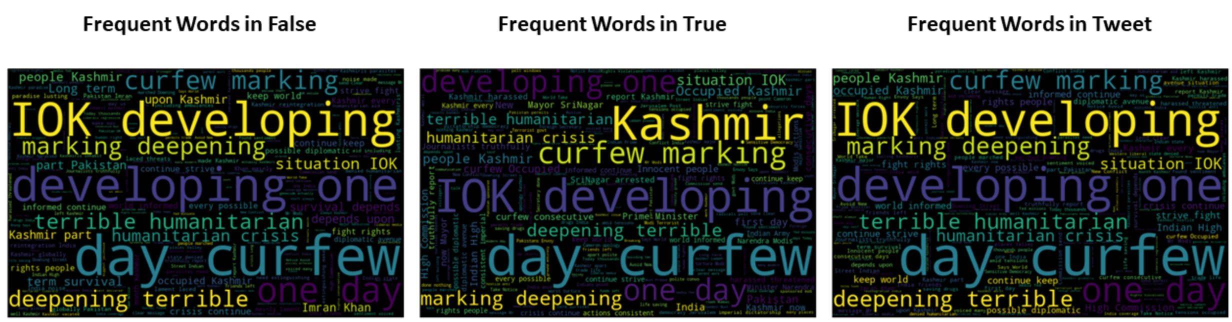

Fig. 4. Comparison of frequent words in fake and real tweets.

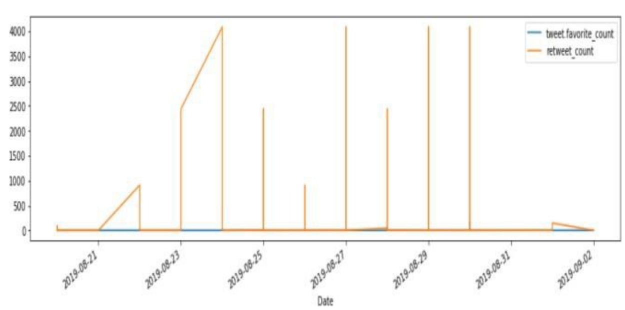

Fig. 5. Most engagement of Twitter user within the selected time span.

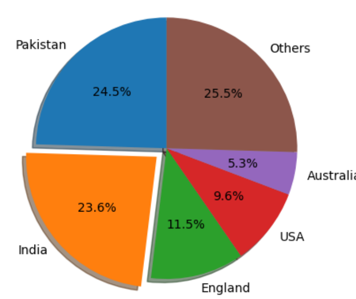

Fig. 6. Tweets percentage based on active Twitter user location.

IV. Discussion

The F1 scores achieved by our best models is 74 per cent which can be attributed primarily to the limited size of our training dataset. Since the tweets on Kashmir Conflict span over few weeks, obtaining a large dataset was not possible, and it was not sufficient to train a DL model. However, our traditional ML models have demonstrated promising results.

Furthermore, our dataset specifically focused on textual

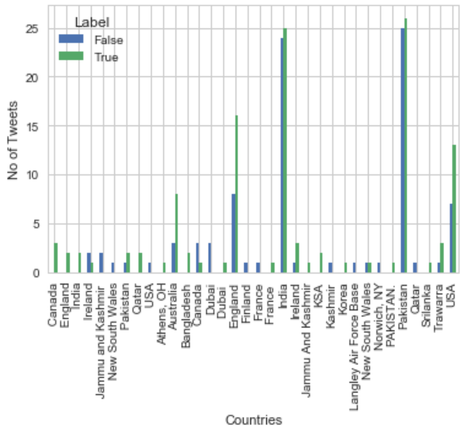

Fig. 7. Comparison of true and fake no. of tweet based on country.

TABLE III. COMPARISON OF DIFFERENT MACHINE LEARNING MODEL BASED ON PERFORMANCE METRICS ON FAKE NEWS DETECTION TASK

Model	F1-Score	Precision	Recall	Accuracy
Random forest	0.74	0.68	0.82	0.82
Naïve Bayes	0.71	0.67	0.76	0.62
SVM	0.73	0.67	0.79	0.79
Logistic regression	0.74	0.68	0.82	0.82
stochastic gradient decent	0.64	0.64	0.63	0.69

data. It is worth noting that existing online fact-checkers, primarily concentrate on fact-checking statements pertaining to American politics [5]. Hence, verifying credibility of real-time news content presented a challenge. Consequently, the task of maintaining an available dataset for our research proved to be challenging, resulting in a decrease in performance when using the "Kashmir Dataset" with a relatively smaller number of instances. Despite these limitations, our research demonstrates the potential of ML techniques in detecting fake news, especially given the constraints of the dataset. We believe that with larger and diverse datasets, our models could achieve higher accuracy and F1 scores.

When data is small, traditional ML has several advantages over deep learning approaches, which typically require large annotated datasets and extensive computational resources. First, it performs well with an F1-score of 74% and accuracy of 82% on a small, event-specific Twitter dataset that represents focusing events. Secondly, our system achieves interpretability and robustness, which cannot be obtained by many end-to-end neural models given that we combine textual features with user profile attributes (e.g., follower ratio, re-tweet count, and verification status). These advantages are reflected in recent comparative studies [61], that compared classical ML algorithms (RF, LR, etc.) to transformer-based models on several real-world datasets and concluded that the traditional ones are still competitive. Likewise, the survey [58] of ML and DL architectures found that Logistic Regression exceeded transformer models in limited environments. Our results confirm these findings because it has been proved that traditional ML is frequently more feasible, more resource-friendly, and interpretable in low-resource, high-impact settings.

V. CONCLUSION

We have developed a system that uses traditional ML techniques to automatically detect and assess the credibility of fake news on Twitter using a comparatively smaller dataset. Our system achieved the highest accuracy of 82 per cent, with RF and LR classifiers. Our study is a contribution to the combating of fake news because we offer a strong system of identification and evaluation of news credibility that gives impressive outcomes using a small dataset. Our experimental results affirm the credibility of traditional ML techniques to analyse "focusing events". Moreover, classification of the most prevalent words in true and false news tweets is useful in highlighting the linguistic patterns that can aid in detecting fake news.

These results indicate the success of our method in real-time, event-specific detection of misinformation, particularly, in the situation when data is scarce. The accuracy and interpretability were achieved by the combination of user profile features and linguistic patterns. Our evaluation does not only provide powerful predictive performance but also adds practical knowledge of how to design lightweight detection systems that can be applicable to political crises and other rapidly changing environments. The framework can be incorporated into early warning systems by journalists, fact-checking organizations, or policy response teams to track misinformation in real-time during political or crisis events. Theoretically, it helps to fill the gap in the under-researched field of fake news detection in focusing events, proving that even in the conditions of limited data, the standard ML models can be effective.

Although the results are promising, this research has limitations. The dataset was small and narrow, only concerned with English tweets during a certain time frame. Such limitations can influence the applicability of the model to more extensive or multilingual fake news detection procedures. Future work will focus on expanding the dataset, incorporating additional feature selection methods, and diversifying language coverage. Additionally, employing different DL techniques, such as fine-tuning a pretrained model can potentially improve the results.

REFERENCES

[1] Seffen, B. and Patt, A., 'A historical turning point? Early evidence on how the Russia-Ukraine war changes public support for clean energy policies." Energy Research & Social Science. vol. 91, pp. 102758, 2022.

[2] Alper Darendeli ,Aixin Sun ,Wee Peng Tay . The geography of corporate fake news. Plos One, 2024.

[3] Li, Hui Ru and Hu, Yan Rong and Liu, Hong Jiu. A novel LASSO-ATT-LSTM model of stock price prediction based on multi-source heterogeneous data. Journal of Intelligent & Fuzzy Systems 1–11, 2023.

[4] Simko, Jakub and Racsko, Patrik and Tomlein, Matus and Hanakova, Martina and Moro, Robert and Bielikova, Maria. (2021). A study of fake news reading and annotating in social media context. New Review of Hypermedia and Multimedia (pp. 97–127), 2021.

[5] Zhou, Xinyi and Zafarani, Reza. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR) (pp. 1–40), 2020.

[6] Sabatier, Paul A. An advocacy coalition framework of policy change and the role of policy-oriented learning therein, 129–168, 1988.

[7] Naeem, Salman Bin and Bhatti, Rubina and Khan, Aqsa. An exploration of how fake news is taking over social media and putting public health at risk. Health Information & Libraries Journal (pp. 143–149), 2021.

[8] Ali, Rizwan and Mustafa, Usman. Kashmir dispute: Emerging complexities after abrogation of Article 370. Journal of Humanities, Social and Management Sciences (JHSMS) (pp. 13–25), 2021.

[9] Conroy, Nadia K and Rubin, Victoria L and Chen, Yimin. Automatic deception detection: Methods for finding fake news. Proceedings of the association for information science and technology (pp. 1–4), 2015.

[10] Buntain, Cody and Golbeck, Jennifer. Automatically identifying fake news in popular twitter threads. 017 IEEE international conference on smart cloud (smartCloud), 208–215, 2017.

[11] Torabi Asr, Fatemeh and Taboada, Maite. Big Data and quality data for fake news and misinformation detection. Big Data and Society. (pp. 127–132). SAGE Publications Sage UK: London, England, 2019.

[12] Girgis, Sherry and Amer, Eslam and Gadallah, Mahmoud. Deep learning algorithms for detecting fake news in online text. In 2018 13th international conference on computer engineering and systems (ICCES) (pp. 93–97), 2018.

[13] Xiao, Cao and Freeman, David Mandell and Hwa, Theodore. Detecting clusters of fake accounts in online social networks. In Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security (pp.91–101), 2014.

[14] Khaled, Sarah and El-Tazi, Neamat and Mokhtar, Hoda MO. Detecting fake accounts on social media. 2018 IEEE international conference on big data (big data), (3672-3681), 2018.

[15] Shu, Kai and Dumais, Susan and Awadallah, Ahmed Hassan and Liu, Huan. Detecting fake news with weak social supervision. IEEE Intelligent Systems, 96–103, 2020.

[16] Benevenuto, Fabricio and Magno, Gabriel and Rodrigues, Tiago and Almeida, Virgilio. Detecting spammers on twitter. Collaboration, electronic messaging, anti-abuse and spam conference (CEAS), 2010.

[17] Ahmed, Hadeer and Traore, Issa and Saad, Sherif. Detection of online fake news using n-gram analysis and machine learning techniques. Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1, 2017.

[18] Gabel, Sabrina and Reichert, Lilian and Reuter, Christian. Discussing conflict in social media: The use of Twitter in the Jammu and Kashmir conflict. Media, War and Conflict (pp. 504–529), 2022.

[19] Kumar, Srijan and West, Robert and Leskovec, Jure. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In Proceedings of the 25th international conference on World Wide Web, 2016.

[20] Jain, Akshay and Kasbe, Amey. Fake news detection. 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2018.

[21] Manzoor, Syed Ishfaq and Singla, Jimmy. Fake news detection using machine learning approaches: A systematic review. 2019 3rd international conference on trends in electronics and informatics (ICOEI), 230–234, 2019.

[22] Granik, Mykhailo and Mesyura, Volodymyr. Fake news detection using naive Bayes classifier. In 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON), 2017.

[23] Dey, Amitabha and Rafi, Rafsan Zani and Parash, Shahriar Hasan and Arko, Sauvik Kundu and Chakrabarty, Amitabha. Fake news pattern recognition using linguistic analysis. In 2018 Joint 7th International Conference on Informatics, Electronics and Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 305–309), 2018.

[24] Di Domenico, Giandomenico and Sit, Jason and Ishizaka, Alessio and Nunan, Daniel. Fake news, social media and marketing: A systematic review. Journal of Business Research (pp. 329–341), 2021.

[25] Juan R. Martinez-Rico ,Lourdes Araujo,Juan Martinez-Romo. Building a framework for fake news detection in the health domain. Plos one, 2024.

[26] Shu, Kai and Mahudeswaran, Deepak and Wang, Suhang and Lee, Dongwon and Liu, Huan. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big data,Mary Ann Liebert, Inc., publishers (pp. 171–188), 2020.

[27] Kumar, Srijan and Shah, Neil. False information on web and social media: A survey. arXiv preprint arXiv:1804.08559, 2018.

[28] Seo, Youngkyung and Seo, Deokjin and Jeong, Chang-Sung. FaNDeR: fake news detection model using media reliability. TENCON 2018-2018 IEEE Region 10 Conference (pp. 1834-1838), 2018.

[29] Khan, Ajmal and Khan, Azmat. From Burhan Wani to Abhi Nandan: A Comparative Analysis of India and Pakistan's Newspaper Coverage of Kashmir Conflict. NUST Journal of International Peace and Stability, 2020.

[30] Mahmood, Z., Daud, A., & Abbasi, R. A. (2021). Using machine learning techniques for rising star prediction in basketball. Knowledge-Based Systems, 211, 106506.

[31] Parikh, Shivam B and Atrey, Pradeep K. Media-rich fake news detection: A survey. 2018 IEEE conference on multimedia information processing and retrieval (MIPR) (pp. 436–441), 2018.

[32] Vieweg, Sarah and Hughes, Amanda L and Starbird, Kate and Palen, Leysia. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1079–1088), 2010.

[33] Fernandez, Miriam and Alani, Harith. Online misinformation: Challenges and future directions. Companion Proceedings of the The Web Conference 2018 (pp. 595–602), 2018.

[34] Samonte, Mary Jane C. Polarity analysis of editorial articles towards fake news detection. In Proceedings of the 2018 1st International Conference on Internet and e-Business (pp. 108–112), 2018.

[35] Castillo, Carlos and Mendoza, Marcelo and Poblete, Barbara. Predicting information credibility in time-sensitive social media. Internet Research, 2013.

[36] Clark, Eric M and Williams, Jake Ryland and Jones, Chris A and Galbraith, Richard A and Danforth, Christopher M and Dodds, Peter Sheridan. Sifting robotic from organic text: a natural language approach for detecting automation on Twitter. Journal of computational science, 2016.

[37] A. Daud, N. u. Islam, X. Li, I. Razzak and M. K. Hayat, "Identifying Rising Stars via Supervised Machine Learning," in IEEE Transactions on Computational Social Systems, vol. 10, no. 5, pp. 2374-2384, Oct. 2023, doi: 10.1109/TCSS.2022.3178070.

[38] Behera, Navnita Chadha. The Kashmir conflict: Multiple fault lines. Journal of Asian Security and International Affairs (pp. 41–63), 2016.

[39] Ahmed, Naveed Ur Rehman, Badshah, Afzal, Adeel, Hanan, Tajammul, Ayesha, Duad, Ali, and Alsahfi, Tariq. "Visual Deepfake Detection: Review of Techniques, Tools, Limitations, and Future Prospects," *IEEE Access*, 2024.

[40] Rubin, Victoria L and Conroy, Niall J and Chen, Yimin. Towards news verification: Deception detection methods for news discourse. Hawaii International Conference on System Sciences (pp. 5–8), 2015.

[41] Hussain, Kamal, Saeed, Zafar, Abbasi, Rabeeh, Sindhu, Muddassar, Khattak, Akmal, Arafat, Sachi, Daud, Ali, and Mushtaq, Mubashar. "Towards understanding the role of content-based and contextualized features in detecting abuse on Twitter," *Heliyon*, vol. 10, no. 8, 2024, Elsevier.

[42] Martinez-Gil, Jorge and Chaves-Gonzalez, Jose Manuel. Transfer learning for semantic similarity measures based on symbolic regression, Journal of Intelligent and Fuzzy Systems(pp. 1–13), 2023.

[43] Shu, Kai and Wang, Suhang and Liu, Huan. Understanding user profiles on social media for fake news detection. 2018 IEEE conference on multimedia information processing and retrieval (MIPR) (pp. 430–435), 2018.

[44] Saeed, Zafar, Abbasi, Rabeeh Ayaz, Maqbool, Onaiza, Sadaf, Abida, Razzak, Imran, Daud, Ali, Aljohani, Naif Radi, and Xu, Guandong. "What's happening around the world? A survey and framework on event detection techniques on Twitter," *Journal of Grid Computing*, vol. 17, pp. 279–312, 2019, Springer.

[45] Helmstetter, Stefan and Paulheim, Heiko. Weakly supervised learning for fake news detection on Twitter. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 274–277), 2018.

[46] Abbasi, Rabeeh Ayaz, Maqbool, Onaiza, Mushtaq, Mubashar, Aljohani, Naif R., Daud, Ali, Alowibdi, Jalal S., and Shahzad, Basit. "Saving lives using social media: Analysis of the role of Twitter for personal blood

donation requests and dissemination," *Telematics and Informatics*, vol. 35, no. 4, pp. 892–912, 2018, Elsevier.

[47] Gupta, Aditi and Lamba, Hemank and Kumaraguru, Ponnurangam and Joshi, Anupam. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. Proceedings of the 22nd international conference on World Wide Web (pp. 729–736), 2013.

[48] Banjar, Ameen, Ahmed, Zohair, Daud, Ali, Abbasi, Rabeeh Ayaz, and Dawood, Hussain. "Aspect-Based Sentiment Analysis for Polarity Estimation of Customer Reviews on Twitter," *Computers, Materials & Continua*, vol. 67, no. 2, 2021.

[49] Yasmin, A., Haider, W., Daud, A., & Banjar, A. Machine learning based software effort estimation using development-centric features for crowdsourcing platform. Intelligent Data Analysis, (Preprint), 1-31.

[50] Alowibdi, Jalal S., Alshdadi, Abdulrahman A., Daud, Ali, Dessouky, Mohamed M., and Alhazmi, Essa Ali. "Coronavirus pandemic (COVID-19): Emotional toll analysis on Twitter," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 17, no. 2, pp. 1–21, 2021, IGI Global.

[51] M. Luqman, M. Faheem, W. Y. Ramay, M. K. Saeed and M. B. Ahmad, "Utilizing Ensemble Learning for Detecting Multi-Modal Fake News," in IEEE Access, vol. 12, pp. 15037-15049, 2024, doi: 10.1109/ACCESS.2024.3357661.

[52] Atodiresei, Costel-Sergiu and selea, Alexandru and Iftene, Adrian. Identifying fake news and fake users on Twitter. Procedia Computer Science, Elsevier. (pp. 451–461), 2018.

[53] Gurajala, Supraja and White, Joshua S and Hudson, Brian and Matthews, Jeanna N. Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach. Proceedings of the 2015 international conference on social media and society (pp. 1–7), 2015.

[54] Figueira, Alvaro and Oliveira, Luciana. The current state of fake news: challenges and opportunities. Procedia computer science (pp. 817–825), 2017

[55] Kumar, Mahesh and Kumar, Devender. An efficient gravitational search decision forest approach for fingerprint recognition. Kuwait Journal of Science, 2023.

[56] Saeed, R., Afzal, H., Rauf, S. A., & Iltaf, N. (2023). Detection of offensive language and its severity for low resource language. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(6), 1-27.

[57] Alghamdi, J., Lin, Y., & Luo, S. (2022). A comparative study of machine learning and deep learning techniques for fake news detection. *Information*, 13(12), 576. MDPI.

[58] Binay, A., Binay, A., & Register, J. (2024). Fake news detection: Traditional vs. contemporary machine learning approaches. *Journal of Information & Knowledge Management*, 23(05), 2450075. World Scientific.

[59] Nasser, M., Arshad, N. I., Ali, A., Alhussian, H., Saeed, F., Da'u, A., & Nafea, I. (2025). A systematic review of multimodal fake news detection on social media using deep learning models. *Results in Engineering*, 26, 104752. Elsevier.

[60] Yi, J., Xu, Z., Huang, T., & Yu, P. (2025). Challenges and innovations in LLM-powered fake news detection: A synthesis of approaches and future directions. In *Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security* (pp. 87–93).

[61] Alghamdi, J., Lin, Y., & Luo, S. (2022). A comparative study of machine learning and deep learning techniques for fake news detection. *Information*, 13(12), 576. MDPI.

[62] Hu, B., Mao, Z., & Zhang, Y. (2025). An overview of fake news detection: From a new perspective. *Fundamental Research*, 5(1), 332–346. Elsevier.