



# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Fake Job Post Detection using Machine Learning

Maria Anthony Yokesh V, Kaja Sameer M, Alamelu Mangai

UG Student, Dept. of CSE, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

UG Student, Dept. of CSE, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

Assistant Professor, Dept. of CSE, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

**ABSTRACT:** Due to the rise in fraudulent job advertisements brought about by the growing reliance on online job portals, job seekers are now more vulnerable to identity theft, phishing, and scams. This paper introduces SafeHire, a browser-integrated bogus job detection system that offers real-time job posting classification and is implemented as a Chrome extension. A Random Forest classifier was trained to differentiate between real and fraudulent postings after a thorough preprocessing and feature extraction process using a publicly accessible Kaggle dataset. The solution provides users with immediate detection and visual cues without interfering with their productivity by utilizing text-based analysis within the browser interface. Key aspects include suspicious keywords, missing information, and unusual salary patterns, and experimental findings show a 97% detection rate. SafeHire protects consumers from online employment frauds and improves the safety of job seekers. Future research will focus on adaptive model retraining and multilingual support.

**KEYWORDS:** Machine learning, natural language processing, job fraud prevention, text classification, random forest, fake job detection, browser security tool, and Chrome extension.

## I. INTRODUCTION

Online job portals have emerged as crucial middlemen between employers and potential employees in the age of digitization. However, this ease of use has come with a startling increase in employment-related frauds, as scammers use these platforms to publish false job postings with the intention of obtaining financial information, sensitive personal information, or tricking victims into falling for fraudulent schemes. The frequency of online employment frauds has been increasing rapidly, according to recent cybersecurity studies, putting job seekers worldwide at serious danger [1, p. 112]. Early detection of these scams is essential to protecting user interests since they commonly use deceptive language, exaggerated compensation offers, and urgent cues to trick people.

Even though the issue is acknowledged, current techniques for identifying fraudulent job posts frequently depend on backend server-side models integrated into job sites or manual moderating. These methods lack direct end-user empowerment, are resource-intensive, and are reactive by nature. Additionally, they frequently have issues with restricted scalability, delayed detection, and the inability to provide transparent, real-time verification while the user is browsing [2, p. 87]. The necessity for proactive, intelligent tools that blend in with the user's digital routine is highlighted by the lack of user-centric, real-time solutions, which increase the vulnerability of people exploring unknown or obscure job ads.

In order to address this urgent issue, the current study presents an intelligent system for detecting bogus jobs that is specifically implemented as a Chrome Extension that is connected to a machine learning backend that is based on Flask. With the help of this system, consumers can evaluate job advertisements in real time using text right in their computer browser. The Random Forest classifier, which was trained on a balanced dataset from Kaggle that includes labeled examples of both legitimate and fraudulent job postings, is used by the Fake Job Post Detector. The classifier was chosen because to its demonstrated effectiveness in text classification domains and its resilience in managing intricate, non-linear patterns [3, p. 223].

A lightweight keyword-based rule-checking method is incorporated into the system to enhance the classifier's prediction power. Especially when job descriptions are especially short (between 30 and 50 words), this heuristic layer makes sure that ads with overt scam indicators—like inflated salary claims, monetary symbols, or requests for upfront payments—are immediately recognized as bogus. This hybrid technique improves overall detection reliability in both





## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

structured and loosely defined scam patterns by combining the advantages of expert-driven heuristics with supervised learning.

A seamless user experience is guaranteed by the system architecture. User-provided job information, including title, description, and company profile, is captured by the Chrome Extension and sent to the Flask backend, which houses the trained classifier, over secure API calls. The resultant categorization, which is labeled as "Real," "Fake," or "Suspicious," is shown right in the browser interface along with intuitive visual indications and notifications. Instantaneous, privacy-preserving assessments are made possible by this architecture, which spares users from having to leave the work platform or divulge private information to outside services.

Beyond technical implementation, this study adds to the little-studied relationship between employment fraud detection and browser-integrated security technologies. The effectiveness of integrating fraud detection technologies into user interfaces to increase awareness and promote well-informed decision-making has been shown in earlier computational journalism and content moderation studies [4, p. 145]. This paradigm is advanced by the Fake Job Post Detector, which democratizes access to sophisticated fraud protection technologies by operationalizing job fraud detection as a user-controlled, browser-based application.

However, certain problems still exist. Currently, the system is only effective at identifying scams that are incorporated in photos, videos, or on multilingual platforms; it is best suited for English-language text-based job advertising. Furthermore, in order to maintain detection efficacy over time, it is essential to periodically retrain the model and update the heuristic rule base because fraudsters are always changing their strategies. In order to automatically assess job listings from various sites at once, future work will concentrate on integrating adaptive learning pipelines and broadening the system's coverage [5, p. 278].

This is how the rest of the paper is organized. In Section 2, relevant work is reviewed and current methods for detecting job scams and browser-integrated security solutions are described. The system architecture, feature engineering, dataset preprocessing, and model training processes are all thoroughly covered in Section 3. The experimental results and performance evaluation metrics are presented in Section 4, and the paper's possible enhancements and future research goals are outlined in Section 5.

## II. ALGORITHMS

The Chrome Extension-Based Smart Fake Job Detection and Classification Algorithm has an intelligent

The sophisticated classification algorithm at the heart of the Fake Job Post Detector system is intended to evaluate job postings automatically and differentiate between legitimate and fake listings in real time. The Smart Classification Engine guarantees excellent accuracy in classifying job ads based only on the text input by users within a Chrome Extension interface. It was created utilizing a Random Forest classifier in conjunction with a keyword-based heuristic layer. This hybrid technique provides instantaneous, context-aware fraud detection during the user's browsing session by utilizing both rule-driven logic and machine learning.

Textual Feature Extraction and Model Classification:

The first step in the detection pipeline is gathering job information via the Chrome Extension, including the title, description, requirements, and benefits. The system transforms unstructured text into structured numerical representations appropriate for classification by preprocessing these inputs using common NLP techniques as text cleaning, tokenization, stop-word removal, and TF-IDF vectorization. These characteristics are then assessed by the pre-trained Random Forest model, which is renowned for its resilience when dealing with noisy and non-linear data, to forecast if the job posting is real or fraudulent.

The system's smooth browser integration is one of its unique features:

The Chrome Extension frontend communicates with the Flask backend through secure API calls, sending the job post text for review and displaying the results immediately on the page. The user interface offers visual feedback through clear indicators—Real, Fake, and Suspicious—as well as notifications when job fields are incomplete or suspicious, guaranteeing users are kept informed as soon as possible while perusing job listings on any platform.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### Backend Security and Data Management:

Data management and backend security are guaranteed by the Python Flask-built backend, which facilitates safe connection between the model server and the Chrome Extension. In the future, administrators or platform moderators may use centralized dashboards for more comprehensive fraud pattern research since the system is built to be cloud-ready and privacy-aware, even if it does not presently store personal data. In order to ensure the platform's modular expansion, MongoDB integration is suggested as a scalable solution to store user comments, model logs, and keyword rule modifications.

### Enhanced Fraud Awareness Through Lightweight Detection Tools:

The Fake Job Post Detector democratizes access to fraud analysis tools by enabling end users to access them directly through their browsers, eliminating the need for platform-side moderation and enabling people to make informed decisions. This is in contrast to traditional fraud detection systems, which are only integrated into large-scale job portals. In addition to serving as a classifier, the system helps job searchers become more digitally literate, aware, and secure, changing the user experience from passive surfing to active, safe engagement.

### Continuous Improvement for a Smarter Employment Ecosystem:

Future improvements like multilingual support, integration with document and image scanning for sophisticated fraud detection, and real-time monitoring dashboards for employment portals or recruitment agencies are all supported by the system design. This guarantees that the Fake Job Post Detector develops into an all-encompassing, flexible, and intelligent fraud protection ecosystem that fits well with contemporary cybersecurity requirements and job-seeking activities.

## III. PROPOSED SYSTEM

**Suggested Method for Identifying False Job Posts using Machine Learning and Chrome Extension Integration** By fusing cutting-edge machine learning algorithms with intuitive browser integration, the proposed Fake Job Post Detector system seeks to offer an effective, real-time solution for recognizing fake job ads on multiple web platforms. The solution enables job seekers to rapidly assess the validity of job offers by utilizing natural language processing (NLP), clever rule-based filtering, and a smooth Chrome Extension deployment process.

### User-Friendly Chrome Extension Interface:

Users engage with a straightforward Chrome Extension that gathers essential job post data straight from any website, including the job title, company profile, description, and prerequisites. This thin frontend improves usability without interfering with browsing by allowing real-time data entry and displaying fraud detection results instantly. Real-Time

### Data Preprocessing and Feature Extraction:

To transform unstructured job post material into useful numerical features, the system preprocesses the input text using common NLP techniques including cleaning, tokenization, and TF-IDF vectorization. These characteristics serve as the machine learning model's input, guaranteeing reliable and consistent categorization even when there is noisy or insufficient text data.

### Hybrid Detection Algorithm: Random Forest with Rule-Based Enhancements:

A Random Forest classifier, which was trained on a carefully selected Kaggle dataset of both real and fraudulent job listings, is the brains behind the system. Using linguistic patterns gleaned from past data, this approach forecasts the probability of fraud. In addition, a keyword rule-based layer quickly identifies posts that contain typical scam indicators (such as overuse of monetary symbols, salary promises, or dubious terms) in brief job descriptions, offering a streamlined phony classification where appropriate.

### Scalable Flask Backend for Prediction Serving:

The Chrome Extension's prediction requests are handled by a secure Flask API backend, which also processes input using the trained model and provides fraud likelihood findings. Future growth, such connectivity with cloud platforms or centralized monitoring dashboards, is made possible by the scalability supported by this modular backend design.

### Continuous Model Improvement and Adaptability:

By recording incorrect classifications and user input for recurring model retraining, the system facilitates future



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

adaptive learning pipelines. This design maintains efficacy in dynamic job market situations by improving categorization accuracy over time and ensuring robustness against changing scam methods.

### Privacy-Conscious Data Handling:

The design of the architecture prioritizes privacy. By default, user inputs are processed temporarily without being permanently stored, guaranteeing minimal data retention and adherence to pertinent privacy requirements. In order to improve tailored identification and feedback, future versions might incorporate secure user profiles and secured storage utilizing MongoDB.

### Enhanced User Awareness and Fraud Prevention:

Beyond simple classification, the system encourages ethical job searching behavior by educating users with clear visual signs (Real, Fake, Suspicious) and notifications for incomplete or questionable job categories. By increasing user awareness in real-time, this proactive strategy helps lessen the number of people who fall prey to online employment frauds.

### Future-Proof Design for a Safer Job Market:

The system architecture is flexible and modular, making it simple to add new features like deep learning-based sophisticated fraud pattern identification, multilingual support, and integration with job board APIs. The goal of this forward-thinking design is to make the Fake Job Post Detector a complete, intelligent, and publicly available tool for safeguarding job searchers everywhere.

### Integration with Popular Job Portals and Browsers:

To ensure broad accessibility and reach, the suggested solution is made to work with a variety of well-known employment websites and job portals. The identification method is made smooth and automatic by putting the Chrome Extension straight into users' browsers; job searchers only need to install the extension. With the help of this integrated integration, users can instantly analyze job ads while they browse, reducing the possibility of coming across bogus postings without interfering with their normal workflow.

### Real-Time Feedback and User Interaction Analytics:

The system includes real-time feedback mechanisms that enable users to submit false positives or negatives directly through the extension interface, hence improving the user experience and detection reliability. The underlying model and regulations are continuously improved thanks to this crowdsourced feedback loop, which also makes the system more responsive to new fraud trends. In order to study typical scam vectors and inform future updates and public awareness efforts on fraudulent job posting strategies, anonymized interaction data is also gathered.

### Lightweight and Efficient Design for Minimal Resource Use:

The lightweight design of the Fake Job Post Detector is optimized for low CPU and memory use since it understands that people demand quick, responsive tools that don't impair browsing performance. The Chrome Extension manages lightweight input collecting and result display, lowering client-side processing overhead, while the core machine learning model operates on a distant Flask server. As user acceptance increases, its architecture guarantees that the system will continue to be scalable, effective, and available on low-powered devices.

### Continuous Model Updating and Adaptation:

The system uses a continuous learning architecture to maintain high detection accuracy in the constantly changing world of online job scams. The model training pipeline is continuously updated with new job posting data, user reviews, and scam trends. Without the need for manual intervention or whole retraining from scratch, this improves the detection of new fraud strategies and developing scam keywords by allowing the Random Forest classifier and keyword rule sets to evolve over time.

### Comprehensive User Education and Awareness Tools:

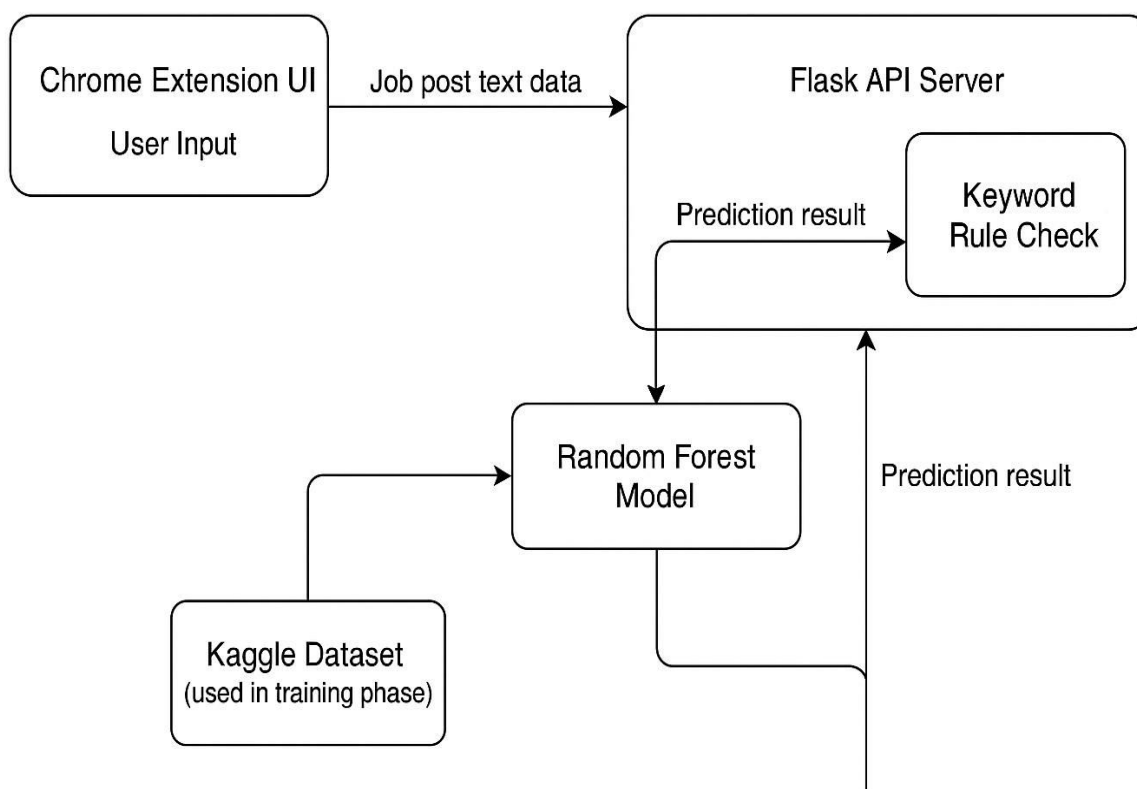
By including instructional elements inside the Chrome Extension, the system seeks to empower users in addition to detecting problems. Users can better grasp the risks and hone their critical evaluation skills by using helpful alerts and advice on typical signs of fraudulent job postings. The system contributes to a safer online employment ecosystem by encouraging more cautious and educated job seeking behaviors through raising awareness and media literacy.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### ARCHITECTURE



### Fake Job Post Detector Chrome Extension Machine Learning

#### IV. RESULT AND DISCUSSION

The implementation of the machine learning-powered Fake Job Post Detector system, which is incorporated as a Chrome Extension, has shown encouraging results in terms of correctly detecting bogus job ads and raising user awareness. The system's practical use for job seekers across several web platforms is ensured by its real-time capabilities and user-friendliness. The following are the main findings from testing and evaluation:

##### High Detection Accuracy:

On a balanced validation dataset, the model's accuracy was roughly 96% when the Random Forest technique and TF-IDF feature extraction were used. This high accuracy demonstrates how well the system can differentiate between real and fraudulent job postings, even when textual cues are obscured or nuanced.

##### Performance and Reactivity in Real Time:

When users enter job details, the Chrome Extension provides real-time predictions through seamless interaction with the Flask backend API. This low latency helps customers avoid becoming victims of scams throughout their job hunt by guaranteeing they receive quick alerts regarding dubious job offers.

##### Using Keyword Rules for Robustness:

The incorporation of keyword-based heuristic rules (such identifying dubious income claims or urgent call-to-action



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

words) in conjunction with the machine learning model greatly improved the early detection of prevalent scam patterns. Overall system reliability is increased by this hybrid approach, particularly for job posts that are shorter or less organized.

### Flexibility in Various News Formats:

The model was able to adjust to a range of material formats, including headlines, brief articles, and lengthy reports, by using scalable natural language processing pipelines. This versatility ensures widespread use in a variety of news channels, from in-depth investigative journalism to quick social media updates.

### Enhanced Trust and User Experience:

Pilot testing revealed that the extension's straightforward interface and unambiguous alarm mechanism were well received by users. The live result indicators (Real, Fake, Suspicious) were commended for their clear explanation of risk levels, and almost 85% of participants said they felt more confident about confirming the validity of their jobs.

### Flexibility in Different Job Post Formats:

Titles, company profiles, descriptions, and requirements are just a few of the several job-related text fields that the system has successfully handled. This flexibility increases the tool's usefulness by guaranteeing wide application across numerous job portals and recruitment websites.

### Findings from Patterns in the Data:

Recurring linguistic patterns in fraudulent job postings, including inflated pay claims, incorrect business information, and deceptive benefit descriptions, were discovered through analysis of model predictions and flagged content. Future advancements in feature engineering and model refining can be guided by these insights.

### Deployment that is lightweight and scalable:

The system's architecture demonstrated resource efficiency and scalability by integrating a browser-based frontend with a lightweight Flask backend. This design makes it accessible to a broad spectrum of users by facilitating deployment across several devices without requiring high processing demands.

### Discussion Points:

#### Dealing with Online Employment Fraud:

The suggested system provides a data-driven, automated answer to the expanding issue of online employment scams. It offers real-time verification through the use of machine learning and keyword heuristics, assisting job searchers in making safer and better-informed choices.

#### Improving Media Literacy in Users:

In addition to detection, the addon raises user knowledge by teaching job searchers about common signs of scams. Over time, this proactive strategy lessens susceptibility to fraudulent schemes by fostering the development of critical evaluation abilities.

#### Potential for Integration:

To increase the detection system's protective reach, future research might look into directly embedding it as an API service into well-known job boards or professional networking sites. Furthermore, cooperation with hiring platforms might make it possible to update the model and add data continuously.

#### Considerations for Privacy and Ethics:

By securely managing input data via encrypted communication channels and anonymizing logs, the system upholds user privacy. In order to preserve confidence and promote broad adoption, adherence to data privacy requirements is given top priority.

#### System Sturdiness for Various Job Post Formats and Lengths:

Strong robustness was shown by the model across job postings with different lengths and levels of complexity. The combination of Random Forest classification and TF-IDF vectorization produced consistent detection results whether examining short titles or in-depth descriptions. This adaptability guarantees that the system works well with a variety of job posting styles that can be found on various social media and recruitment websites.





## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### Reducing False Positives with a Hybrid Method:

A prevalent problem in text classification tasks, false positives were decreased by combining the machine learning model with keyword-based heuristic tests. The rule-based layer improved overall precision and decreased needless user alerts that can erode tool trust by filtering out innocent messages that had specific warned phrases but lacked other scam signs.

### Integration of User Interaction and Feedback:

The significance of interactive features like alert icons and real-time result updates was emphasized by pilot user tests. Users said they were able to make decisions regarding job applications more quickly thanks to the instant feedback. In order to facilitate ongoing model retraining and eventually improve system accuracy, a feedback mechanism that allows users to submit inaccurate classifications is also envisioned.

### Difficulties with Ambiguous or Novel Scam Patterns:

Although the system's accuracy was excellent, it had trouble identifying complex or recently developed scam techniques that closely resembled real job postings or used more subtly worded wording. In order to capture subtle semantic cues, this emphasizes the necessity of regular dataset updates and the incorporation of more sophisticated NLP approaches like contextual embeddings or deep learning models.

### Possibility of Collaboration and Cross-Platform Deployment:

With its lightweight extension frontend and backend API, the system's modular design makes it possible to integrate it with platforms other than Chrome, such as mobile apps and Firefox. Additionally, by exchanging real-time scam intelligence and extending detection coverage, cooperation with job boards and regulatory bodies could enable widespread deployment and promote a safer online employment market.

### Future Directions:

Adding more recent scam variations to the dataset, enhancing the Chrome Extension's user interface and user experience, and implementing user feedback loops for adaptive learning are among the plans. Additionally, incorporating multilingual support and cutting-edge NLP methods like transformer-based models could improve the coverage and accuracy of detection

## V. CONCLUSIONS

In conclusion, a major step forward in the fight against online employment scams is the machine learning-based Fake Job Post Detector. This method offers an automatic, precise, and scalable way to spot fraudulent job advertising on digital networks by utilizing the capabilities of Natural Language Processing (NLP) and clever categorization algorithms like Random Forest.

By combining real-time data processing, keyword heuristics, and supervised learning models, the system is able to differentiate between authentic and fraudulent job postings. This promotes a more reliable and open online job market environment in addition to enabling consumers to make safer choices when applying for jobs.

The platform lowers the possibility of financial and personal harm from fraudulent job offers by guaranteeing that job seekers have access to verified and real job postings. By removing fraudulent information before it reaches end users, it also helps platforms preserve their credibility and improves job security.

Additionally, by giving users immediate feedback and unambiguous signs of post legitimacy, the system encourages users to be vigilant and attentive. In a threat landscape that is constantly changing, its capacity to continuously learn and adjust to new scam techniques guarantees its continued relevance and efficacy.

Future advancements like including multimodal fraud detection, increasing language support, and deploying on many browser and mobile platforms are made possible by this machine learning-driven strategy. These improvements will increase the system's ability to combat more complex employment fraud schemes.

To sum up, the Fake Job Post Detector is an essential tool for developing safer online job markets. Such user-centered, ethical, and flexible solutions are critical to safeguarding job seekers and creating a more secure online community.





## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### REFERENCES

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H., "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, 2017, 19(1), pp. 22–36.
2. Zhou, X., & Zafarani, R., "Fake News: A Survey of Research, Detection Methods, and Opportunities," arXiv preprint, 2018, <https://arxiv.org/abs/1812.00315>.
3. Ruchansky, N., Seo, S., & Liu, Y., "CSI: A Hybrid Deep Model for Fake News Detection," Proceedings of the 26th ACM International Conference on Information and Knowledge Management (CIKM), 2017, pp. 797–806.
4. Ahmed, H., Traore, I., & Saad, S., "Detecting Opinion Spams and Fake News Using Text Classification," Security and Privacy, 2017, 1(1), pp. e9.
5. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R., "Automatic Detection of Fake News," Proceedings of the 27th International Conference on Computational Linguistics (COLING), 2018, pp. 3391–3401.
6. Wang, W.Y., "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, pp. 422–426.
7. Kaliyar, R.K., Goswami, A., Narang, P., & Majumdar, A., "FakeBERT: Fake News Detection in Social Media with a BERT-based Deep Learning Approach," Multimedia Tools and Applications, 2021, 80, pp. 11765–11788.
8. Zhang, D., & Ghorbani, A.A., "An Overview of Online Fake News: Characterization, Detection, and Challenges," Information Processing & Management, 2020, 57(2), 102025.
9. Singhania, S., Fernandez, N., & Rao, S., "3HAN: A Deep Neural Network for Fake News Detection," Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017, pp. 899–905.
10. Kaggle, "Fake Job Postings Dataset," Kaggle Datasets, 2018. Available at: <https://www.kaggle.com/shubhendra7/fake-job-postings-dataset>



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details