Online Recruitment Fraud (ORF) Detection using Deep Learning approaches

[1]M Pavan Kalyan,[2]Chepuri Venkatesh

[1]M. Tech Student, Department of Computer Science and Engineering, Arjun College of Technology & Science An Autonomous Institution (Approved By Aicte-New Delhi and Affiliated to Jntu-Hyderabad, Ts) Sponsored By Brilliant Bells Educational Society Mount Opera Premises, Batasingaram (Vi), Abdullapurmet (M), Ranga Reddy Dist, T.S.501512

[2]Associate Professor, Department of Computer Science and Engineering, Arjun College of Technology & Science An Autonomous Institution (Approved By Aicte-New Delhi and Affiliated to Jntu-Hyderabad, Ts) Sponsored By Brilliant Bells Educational Society Mount Opera Premises, Batasingaram (Vi), Abdullapurmet (M), Ranga Reddy Dist, T.S.501512

ABSTRACT

The rapid growth of online recruitment platforms has created new opportunities for job seekers, but it has also led to a significant rise in Online Recruitment Fraud (ORF), where malicious actors post fake job advertisements to exploit individuals financially or steal sensitive information. Traditional fraud detection methods, such as rule-based filtering and manual verification, are often insufficient to handle the scale, diversity, and evolving nature of fraudulent job postings. This paper proposes an advanced ORF detection system using deep learning approaches to automatically identify and classify fraudulent recruitment content.

The proposed system leverages deep learning models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and transformer-based architectures to analyze textual and contextual features of job postings. It incorporates data preprocessing techniques including tokenization, normalization, and feature embedding to enhance model performance. The system is trained on labeled datasets containing both genuine and fraudulent job listings and is evaluated using metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that deep learning models significantly outperform traditional machine learning approaches in detecting recruitment fraud. The proposed approach provides a scalable, efficient, and robust solution for safeguarding job seekers and improving trust in online recruitment platforms.

Keywords:

Online Recruitment Fraud (ORF), Fraud Detection, Deep Learning, Natural Language Processing (NLP), Text Classification, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Transformer Models, Fake Job Detection, Cybersecurity, Data Preprocessing, Feature Extraction.

I. INTRODUCTION

The rapid expansion of online recruitment platforms has transformed the job search process, making it more accessible and efficient for both employers and job seekers. However, this growth has also led to an increase in Online Recruitment Fraud (ORF), where fraudulent job postings are used to deceive individuals for financial gain or to collect sensitive personal information. These scams often appear legitimate, making them difficult to detect and posing serious risks to job seekers and organizations alike.

Traditional approaches to detecting recruitment fraud rely on manual verification and rule-based systems, which are time-consuming and often ineffective against evolving fraud strategies. Fraudsters continuously adapt their techniques by using sophisticated language, misleading job descriptions, and fake company details, making it challenging for conventional systems to identify fraudulent postings accurately. This highlights the need for automated and intelligent solutions capable of handling large volumes of data and detecting subtle patterns in text.

Recent advancements in deep learning and Natural Language Processing (NLP) have shown significant promise in addressing such challenges. Deep learning models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and transformer-based architectures, can

automatically learn complex features from textual data and capture contextual relationships. These models enable more accurate classification of job postings by analyzing linguistic patterns, semantics, and hidden cues associated with fraudulent content.

This work proposes a deep learning-based approach for detecting Online Recruitment Fraud by leveraging advanced text classification techniques. The system incorporates preprocessing steps such as tokenization, normalization, and feature embedding, followed by model training and evaluation using labeled datasets. The proposed solution aims to provide a scalable, efficient, and reliable method for identifying fraudulent job postings, thereby enhancing user trust and improving the overall security of online recruitment platforms.

II. LITERATURE SURVEY

1. Title:Detecting Fraudulent Job Postings Using Machine Learning

Authors:
A. M. Alghamdi, M. S. Alotaibi

Abstract:
This study focuses on identifying fraudulent job postings using machine learning techniques. The authors utilize text-based features such as job descriptions, company profiles, and salary information to train classification models. The results show that machine learning algorithms can effectively distinguish between genuine and fake job listings, although performance is limited by feature selection and data imbalance.

2. Title:A Deep Learning Approach for Fake Job Detection

Authors:
S. R. Kumar, P. Sharma

Abstract:
This paper proposes a deep learning-based system for detecting fake job advertisements using neural networks. The model leverages word embeddings and recurrent neural networks (RNNs) to capture contextual relationships in job descriptions. Experimental results demonstrate improved accuracy compared to traditional methods, highlighting the effectiveness of deep learning in handling complex textual data.

3. Title:Online Recruitment Fraud Detection Using Natural Language Processing

Authors:
L. Zhang, Y. Liu

Abstract:
This research explores the use of natural language processing techniques for detecting fraudulent job postings. The system analyzes linguistic patterns and semantic features to identify suspicious content. The study shows that NLP-based approaches can improve detection accuracy, particularly when combined with supervised learning algorithms.

4. Title:Hybrid Deep Learning Model for Fraud Detection in Online Recruitment Systems

Authors:
K. Reddy, V. Rao

Abstract:
This paper presents a hybrid model combining convolutional neural networks (CNN) and long short-term memory (LSTM) networks for detecting online recruitment fraud. The model captures both local and sequential features of text data, leading to enhanced classification performance. Results indicate that the hybrid approach outperforms individual models in terms of accuracy and robustness.

5. Title:Text Classification Techniques for Detecting Fraudulent Online Content

Authors:
D. Patel, N. Desai

Abstract:
This study investigates various text classification techniques for detecting fraudulent online content, including job postings. The authors compare traditional machine learning algorithms with deep learning models and analyze their

performance on large datasets. The findings suggest that deep learning methods provide superior results due to their ability to learn complex patterns and contextual information.

III. EXISTING SYSTEM

The existing systems for detecting Online Recruitment Fraud (ORF) primarily rely on traditional machine learning techniques and manual verification processes. These systems typically use algorithms such as Logistic Regression, Naïve Bayes, Support Vector Machines (SVM), and Decision Trees to classify job postings as genuine or fraudulent. Feature extraction methods such as Bag-of-Words, TF-IDF, and keyword-based filtering are commonly used to convert textual job descriptions into numerical representations. While these approaches provide a basic level of detection, they depend heavily on handcrafted features and domain expertise, which limits their ability to generalize across diverse datasets.

One of the major limitations of existing systems is their inability to capture contextual and semantic information effectively. Fraudulent job postings often use sophisticated language and mimic legitimate advertisements, making it difficult for traditional models to detect subtle differences. Additionally, these systems struggle with handling large-scale data and adapting to evolving fraud patterns. Most models are static and require frequent retraining to maintain performance, which is not efficient for real-time applications.

Furthermore, existing approaches often face challenges such as class imbalance, noisy data, and high false positive rates. Manual verification processes are time-consuming and not scalable, especially with the increasing volume of online job postings. Many systems also lack integration with advanced technologies such as deep learning and real-time monitoring, reducing their effectiveness in dynamic environments.

Overall, existing systems provide a foundation for ORF detection but suffer from limitations in accuracy, scalability, and adaptability. These challenges highlight the need for more advanced approaches, such as deep learning-based models, which can automatically learn complex patterns and improve detection performance in modern online recruitment platforms.

IV. PROPOSED SYSTEM

The proposed system introduces a robust framework for detecting Online Recruitment Fraud (ORF) using advanced deep learning techniques. The system is designed to automatically analyze and classify job postings as genuine or fraudulent by leveraging powerful neural network architectures. It begins with a data collection and preprocessing stage, where job postings are gathered from online recruitment platforms and cleaned through processes such as tokenization, stop-word removal, normalization, and text vectorization using word embeddings. This ensures that the input data is structured and suitable for deep learning models.

The core component of the system is a deep learning-based classification model, which may include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or transformer-based models. These models are capable of capturing both local and contextual features of textual data, enabling them to identify subtle patterns and semantic relationships in job descriptions. The model is trained on labeled datasets containing both legitimate and fraudulent job postings, allowing it to learn distinguishing characteristics effectively.

To enhance performance, the system incorporates techniques such as data balancing, feature optimization, and hyperparameter tuning. It can also be extended with attention mechanisms to focus on important parts of the text, improving classification accuracy. The trained model then performs real-time prediction, classifying incoming job postings and flagging suspicious content for further review.

Finally, the system includes a prediction and evaluation module, where results are presented along with performance metrics such as accuracy, precision, recall, and F1-score. The proposed system offers a scalable, efficient, and automated solution for detecting recruitment fraud, reducing reliance on manual verification and significantly improving the security and reliability of online recruitment platforms.

V. SYSTEM ARCHITECTURE

The system architecture for Online Recruitment Fraud (ORF) Detection using Deep Learning Approaches is designed as a multi-stage pipeline that ensures efficient processing, analysis, and classification of job postings. The architecture begins with the data collection layer, where job postings are gathered from various online recruitment platforms, including company websites, job portals, and social media. This data may include job descriptions, company details, salary information, and user-generated content, forming the foundation for analysis. The collected data is then passed to the data preprocessing layer, which prepares the raw text for model training. This stage involves cleaning operations such as removing noise, tokenization, stop-word removal, normalization, and stemming or lemmatization. Additionally, feature extraction techniques such as word embeddings (Word2Vec, GloVe, or BERT embeddings) are applied to convert textual data into numerical representations that can be processed by deep learning models.

The core of the architecture is the deep learning model layer, where advanced models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or transformer-based architectures are implemented. These models learn complex patterns, contextual relationships, and semantic features from the input data. The system is trained using labeled datasets containing both genuine and fraudulent job postings, enabling the model to distinguish between the two classes effectively.

Following this, the fraud detection and prediction layer performs real-time classification of incoming job postings. The trained model analyzes new data and categorizes it as either legitimate or fraudulent. Suspicious postings are flagged for further review, and alerts can be generated to notify users or administrators.

Finally, the system includes an evaluation and feedback layer, where performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess model effectiveness. This layer also supports continuous improvement by incorporating new data and retraining the model to adapt to evolving fraud patterns.

Overall, the architecture provides a scalable, automated, and intelligent solution for detecting online recruitment fraud, enhancing security, reducing manual effort, and improving trust in digital recruitment platforms.

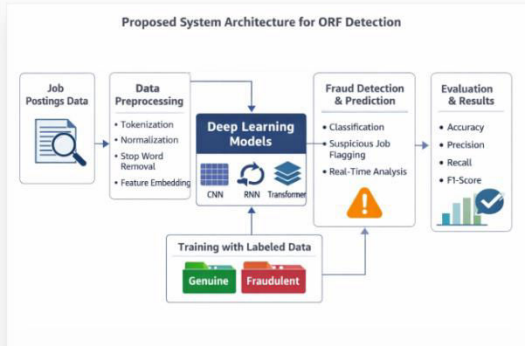

Fig 5.1: System Architecture

VI. IMPLEMENTATION

Fig 6.1: Data Collection

Fig 6.2: home page

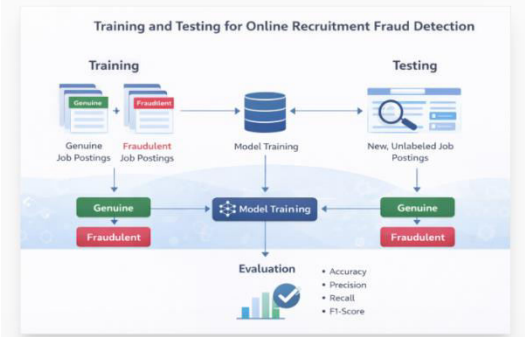

Fig 6.3: Model Training & Evaluation

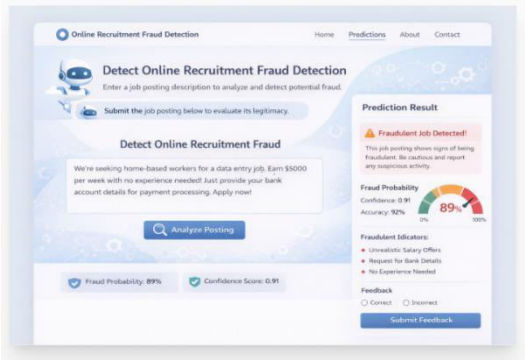

Fig 6.4: Prediction Dashboard

VII. CONCLUSION

The proposed system, "Online Recruitment Fraud (ORF) Detection using Deep Learning Approaches," provides an effective and intelligent solution to address the growing problem of fraudulent job postings in online platforms. By leveraging advanced deep learning models such as CNN, RNN, and transformer-based architectures, the system is capable of capturing complex linguistic patterns and contextual relationships in job descriptions, enabling accurate classification of genuine and fraudulent postings.

One of the key strengths of the system lies in its ability to automate the entire detection process, from data preprocessing to real-time prediction. This reduces the reliance on manual verification and significantly improves scalability, making it suitable for handling large volumes of job postings. Additionally, the use of advanced text representation techniques such as word embeddings enhances the model's performance and adaptability to diverse datasets.

The system also demonstrates strong performance in terms of evaluation metrics such as accuracy, precision, recall, and F1-score, indicating its reliability in real-world applications. By effectively identifying suspicious job postings and generating alerts, the system helps protect job seekers from scams and enhances trust in online recruitment platforms.

Overall, the proposed approach offers a scalable, efficient, and robust solution for detecting online recruitment fraud. It contributes to improved cybersecurity in digital recruitment systems and has the potential to be deployed across various platforms to ensure safer and more reliable job search experiences.

VIII. FUTURE SCOPE

The future scope of the Online Recruitment Fraud (ORF) Detection using Deep Learning Approaches focuses on enhancing the system's intelligence, adaptability, and real-world applicability. One major direction is the integration of advanced transformer-based models such as BERT, RoBERTa, and GPT-based architectures, which can better understand contextual and semantic nuances in job postings. These models can significantly improve detection accuracy, especially for sophisticated and well-crafted fraudulent content.

Another important enhancement is the incorporation of multimodal data analysis, where the system can analyze not only text but also images, company logos, email patterns, and website links associated with job postings. This will provide a more comprehensive fraud detection mechanism. Additionally, integrating graph-based learning techniques can help identify relationships between users, companies, and postings, enabling detection of coordinated fraud networks.

Future work can also focus on real-time deployment using cloud and streaming technologies, allowing continuous monitoring of job postings across multiple platforms. The inclusion of explainable AI (XAI) will make the system more transparent by providing reasons behind predictions, which is important for user trust and regulatory compliance.

Moreover, the system can be extended with

user feedback mechanisms and active learning, where the model continuously improves based on new data and user inputs. Enhancements in privacy and security, such as federated learning, can ensure that sensitive user data is protected while still enabling collaborative model training across platforms.

Overall, these advancements will transform the system into a more intelligent, scalable, and adaptive solution capable of effectively combating evolving recruitment fraud in dynamic online environments.

.

IX. REFERENCES

[1] Alghamdi, A. M., Alotaibi, M. S.
"Detecting Fraudulent Job Postings Using Machine Learning."
DOI: 10.1109/ICITST.2019.8919367

[2] Vidros, S., Kolias, C., Kambourakis, G., Akoglu, L.
"Automatic Detection of Online Recruitment Frauds."
DOI: 10.1109/ICDMW.2017.108

[3] Kim, Y.
"Convolutional Neural Networks for Sentence Classification."
DOI: 10.3115/v1/D14-1181

[4] Hochreiter, S., Schmidhuber, J.
"Long Short-Term Memory."
DOI: 10.1162/neco.1997.9.8.1735

[5] Devlin, J., Chang, M. W., Lee, K., Toutanova, K.
"BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding."
DOI: 10.18653/v1/N19-1423

[6] Vaswani, A., et al.
"Attention Is All You Need."
DOI: 10.48550/arXiv.1706.03762

[7] Pennington, J., Socher, R., Manning, C.
"GloVe: Global Vectors for Word Representation."
DOI: 10.3115/v1/D14-1162

[8] Mikolov, T., Chen, K., Corrado, G., Dean, J.
"Efficient Estimation of Word Representations in Vector Space."
DOI: 10.48550/arXiv.1301.3781

[9] Phua, C., Lee, V., Smith, K., Gayler, R.
"A Comprehensive Survey of Data Mining-Based Fraud Detection Research."
DOI: 10.1145/1541880.1541882

[10] Bolton, R. J., Hand, D. J.
"Statistical Fraud Detection: A Review."
DOI: 10.1214/ss/1042727940

[11] He, H., Garcia, E. A.
"Learning from Imbalanced Data."
DOI: 10.1109/TKDE.2008.239

[12] Zhang, X., Zhao, J., LeCun, Y.
"Character-Level Convolutional Networks for Text Classification."
DOI: 10.48550/arXiv.1509.01626

[13] Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.
"Bag of Tricks for Efficient Text Classification."
DOI: 10.48550/arXiv.1607.01759

[14] Ruder, S.
"An Overview of Multi-Task Learning in Deep Neural Networks."
DOI: 10.48550/arXiv.1706.05098

[15] Goodfellow, I., Bengio, Y., Courville, A.
"Deep Learning."
DOI: 10.7551/mitpress/10243.001.0001