Strategies for Recognizing and Avoiding Fake Job Offers using RNN algorithm

Ms.S.Arularasi[1], Soundaraprasath.D[2], Srikanth.S[3], Jana Praveen.I[4], Sukumar.V[5]
[1],Assistant Professor, Department of CSE, Gnanamani College of Technology, Namakkal,Tamilnadu, India.
[2,3,4,5],UG Students, Department of CSE, Namakkal, Tamilnadu, India.

ABSTRACT- The proliferation of online job portals has revolutionized the employment landscape, providing unprecedented access to job opportunities. However, amidst this convenience lies a lurking threat – the proliferation of fraudulent job postings. These deceptive postings not only deceive job seekers but also tarnish the reputation of legitimate businesses. Detecting fraudulent patterns in job postings is a challenging yet imperative task to safeguard job seekers and maintain the integrity of online recruitment platforms.In this research, we propose a novel approach to unmasking deceptive intent in job postings. Our methodology leverages advanced natural language processing (NLP) techniques, including sentiment analysis, linguistic analysis, and semantic similarity modeling, to discern subtle cues indicative of fraudulent intent. By integrating machine learning algorithms with domain-specific knowledge, our system can effectively identify suspicious patterns that elude traditional rule- based filters.

I.INTRODUCTION

In recent years, the proliferation of online job portals has revolutionized the recruitment landscape, offering both employers and job seekers unprecedented access to a vast array of opportunities and candidates. However, this convenience has also ushered in new challenges, chief among them being the alarming rise in fraudulent job postings. Fraudulent job postings pose a multifaceted threat to both individuals and organizations. For job seekers, falling victim to these scams can result in financial loss, identity theft, and even personal safety risks, as nefarious actors exploit the desperation and vulnerability of those seeking employment. On the other hand, legitimate businesses suffer reputational damage and operational disruptions when their brand is unwittingly associated with fraudulent activities. Addressing this issue is not merely a matter of safeguarding economic interests; it is also crucial for upholding the integrity of the recruitment ecosystem and fostering trust among its participants. As such, there is an urgent need for robust mechanisms to detect and combat fraudulent patterns in job postings.

In response to this imperative, this paper proposes a comprehensive framework for the automated detection of fraudulent patterns in job postings. Leveraging advances in natural language processing (NLP), machine learning (ML), and data mining techniques, our approach aims to identify deceptive elements and anomalous patterns indicative of fraudulent intent. By analyzing textual content, metadata, and contextual cues inherent in job postings, we endeavor to empower stakeholders with the tools necessary to mitigate the risks posed by fraudulent activities.

The remainder of this paper is structured as follows: Section 2 provides a review of related literature, highlighting existing methodologies and gaps in the current state of research. Section 3 outlines the theoretical foundations and conceptual framework underpinning our approach. In Section 4, we detail the methodology employed for data collection, preprocessing, feature engineering, and model development. Section 5 presents experimental results and performance evaluations, demonstrating the efficacy and robustness of our proposed solution. Finally, Section 6 offers insights gleaned from our findings and outlines avenues for future research, concluding with a summary of key contributions and implications for practice.

II. LITERATURE SURVEY

Shawni Datta et al. [6] proposed a methodology to detect the fake jobs scam on the internet using machine learning. Mainly, he used the two types of classifier: single classifier and ensemble classifier, his maximum accuracy was 98.27% in Random Forest classifier. In their research Ibrahim Nasser et al. [9] proposed different machine learning classifiers (Naive Bayes, Support Vector Machine, Decision Tree KNN and Random Forest) to detect the online fake jobs. The highest accuracy was 98.2% for the Random Forest classifier. (R.S. Shishupal, Varsha, S. Mane, V. Singh and

D. Waseker) [11] proposed a methodology to detect the fake jobs by doing communes through speech and message using Natural language Processing (NLP). Accuracy of 96.2% was achieved using this technique. In their research FHA. Shibly et al. [10] proposed two different types of Machine learning algorithms to detect the online fake job vacancies. The accuracy was 93.8% for two class Decision Boosted Trees and 95.4% for two class Decision Forest algorithms. S. Vidros, C.Kolias, G. Kambourakis and L. Akoglu [13] proposed that online job recruitment scam is very dangerous for both companies as well as job seekers. In their research Random Forest was the best algorithm for detection of fake jobs, the accuracy achieved was 91.22%. B. Alghamdi and F. Alharby [1] proposed ensemble classifier approach on the Random Forest algorithm to detect the online fake jobs. He used the Random Forest as an ensemble classifier to detect and classify the cyber attack on the job scams.The accuracy achieved was 97.41% in their research. Pertaining to existing research in the literature, in our paper we employ deep learning techniques to enhance the predictive accuracy of the models and build a robust prediction system.

PROBLEM STATEMENT

The proliferation of online job platforms has led to an increase in fraudulent job postings, posing a significant challenge for both job seekers and employers. Identifying and mitigating fraudulent activities in job postings is crucial

ALGORITHM

Detecting fraudulent patterns in job postings is an important task in today's digital age where online job platforms are susceptible to misuse by scammers. Various algorithms and techniques can be employed to identify such fraudulent activities. Here's an explanation of two commonly used algorithms for this purpose:

Natural Language Processing (NLP) with Machine Learning:Preprocessing: The first step involves preprocessing the text data extracted from job postings. This includes tokenization, removing stopwords, stemming or lemmatization, and vectorization.

Feature Extraction: Features are extracted from the preprocessed text data. These features could include word frequencies, n-grams, sentiment analysis scores, and syntactic or semantic features.

Machine Learning Models: Various machine learning models can be applied to the extracted features. Algorithms like Support Vector Machines (SVM), Random Forest, or Gradient Boosting Machines (GBM) are commonly used. These models are trained on labeled data, where fraudulent and legitimate job postings are already identified.

Representation: In this approach, the job posting data is represented as a graph, where nodes represent entities (such as job titles, companies, skills required) and edges represent relationships between these entities.

to maintain the integrity and trustworthiness of these platforms. However, current approaches often rely on manual inspection or basic rule-based systems, which are time- consuming, inefficient, and prone to human error.

Anomaly Detection: Anomalies or fraudulent patterns are detected by identifying unusual patterns in the graph structure. This could include detecting nodes with unusually high degrees, detecting communities that are disconnected from the main graph, or identifying nodes with unusual behavior compared to their neighbors.

Advantages :

1) Feedback Loop: As new data becomes available and fraudulent techniques evolve, the graph-based approach allows for the continuous updating of the model by incorporating feedback from detected anomalies.

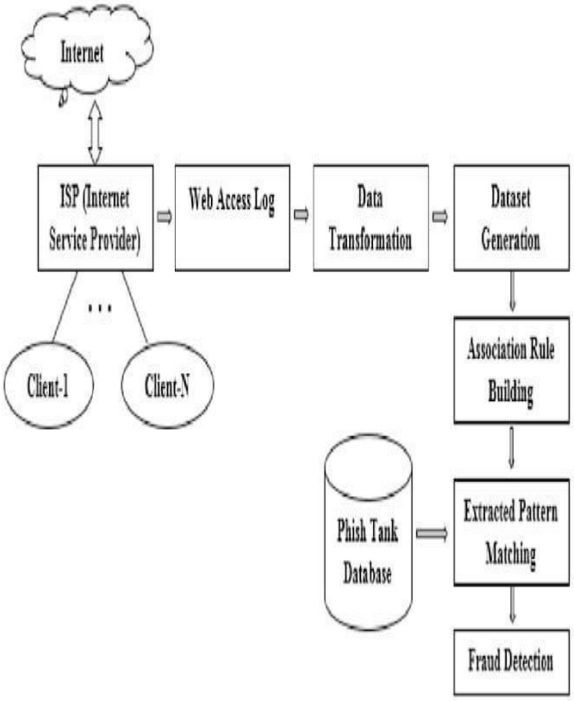
2) Scalability: Graph-based approaches can handle large-scale datasets efficiently, making them suitable for analyzing job posting platforms with a high volume of postings.

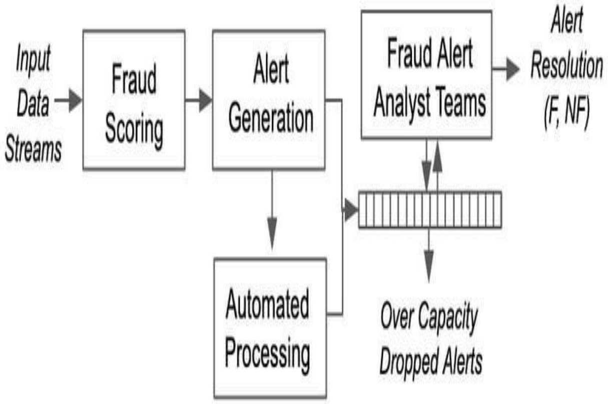
A. Both of these algorithms offer different strengths and can be used either individually or in combination to improve the accuracy and robustness of the fraud detection system for job postings. The choice of algorithm depends on factors such as the nature of the data, available computational resources, and the specific requirements of the application

Block diagram

The system begins with internet access through an Internet Service Provider (ISP), enabling clients to access web content. The web access logs generated by clients' interactions with online platforms are collected and undergo data transformation to prepare them for analysis. This transformed data is then compared against a Phish Tank Database, which contains known instances of fraudulent activities online. Using this comparison, a dataset is generated, capturing potential fraudulent patterns. Next, association rules are built based on the dataset, identifying correlations and patterns indicative of fraudulent behavior. These rules are then applied to the web access logs for pattern matching, extracting potential instances of fraud. Finally, a fraud detection mechanism is activated, flagging or blocking any detected fraudulent activity to protect users from scams and malicious content.

Working
The system functions by initially collecting web access logs from clients via an Internet Service Provider (ISP). These logs undergo preprocessing to ensure uniformity and are then cross-referenced with a Phish Tank Database to establish a dataset highlighting potential indicators of fraudulent activity. Utilizing this dataset, association rules are developed to recognize patterns characteristic of fraudulent behavior. Following this, the web access logs are subjected to pattern matching against these rules to isolate instances suggestive of fraud. Subsequently, a fraud detection mechanism is engaged to promptly identify and respond to any flagged instances of fraudulent activity, such as blocking or notifying users, thus fortifying defenses against scams and deceptive content circulating across online platforms. Through this iterative process of data analysis and pattern recognition, the system effectively detects and mitigates the proliferation of fake job postings and other fraudulent schemes on the internet, enhancing overall security for users and safeguarding against potential financial and reputational harm.Software requirements:





CONCLUSION

In this paper we analyzed and experimented with different machine learning models to identify fake job posting on job portals. We used employer defined linguistic feature to perform our analysis which consisted of various parameters defining a job posted on the portal. We experimented using Random Forest, Support Vector Machines and Bi-Directional LSTMs. The Bi-Directional LSTMs model achieved the highest accuracy of 98.679%, whereas Support Vector Machines achieved the lowest accuracy of 95.773%. The use of employer defined linguistic feature will be advantageous to job portals to automatically identify such job posts and mark them as spam or remove them. The future work should focus on Multilingual Spam Job Posting Detection, so as to have a generalized system for all languages. For future enhancements in fraud detection, several avenues can be explored to further bolster the system's effectiveness and adaptability. Firstly, integrating advanced natural language processing (NLP) techniques can enhance the system's ability to understand and analyze textual content within job postings, allowing for more nuanced detection of fraudulent language patterns and inconsistencies. Additionally, leveraging machine learning algorithms for anomaly detection could help identify previously unseen or evolving fraud tactics by detecting deviations from established patterns or norms in web access logs.

REFERENCES

[1] Alghamdi.B, and Alharby.F, (2019) “An Intelligent Model for Online Recruitment Fraud Detection,” Journal of Information Security, vol. 10, pp.155- 176.

[2] Bansal.S, “ [Real or Fake] Fake JobPosting Prediction. Kaggle,” 9 February2020.

[3] Breiman.L, "Random forests," Machine learning 45.1 pp. 5-32, 2001.

[4] Bhoj.N, et al. “Comparative Analysisof Feature Selection Techniques for Malicious Website Detection in SMOTE Balanced Data.” RS Open Journal on Innovative Communication Technologies, vol. 2, issue 3, pp. 1- 10, 2021, doi:10.46470/03d8ffbd.993cf635.

[5] C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[6] C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[7] C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[8] C.Nagarajan and M.Madheswaran - ‘Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis’- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[9] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. “Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter” Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[10] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya “Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay” Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[11] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash “Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[12] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, “Perfromance Investigation of T-Source Inverter fed with Solar Cell” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[13] C.Nagarajan and M.Madheswaran, “Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation” has been presented in ICTES’08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007

[14] M Suganthi, N Ramesh, “Treatment of water using natural zeolite as membrane filter”, Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[15] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, “Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District”, International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019

[16] Bengio.Y, Ducharme.R, Vincent. P and Jauvin.C, "A neural probabilistic language model," The journal of machine learning research 3, pp.1137-1155, 2003

[17] Datta. S and Bandyopadhyay. S.K, "Fake Job Recruitment Detection Using Machine Learning Approach."

[18] Joyce. S. P, “ 5 Major Types of ScamJobs and Job Scams Online. Job,” 4 October2019.

[19] Lapjaturapit.T, Viriyayudhakom.K and Theeramunkong.T, "MultiCandidate Word Segmentation using Bi-directional LSTM Neural Networks," 2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology