iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73610

Smart Screening: Using Machine Learning to Expose Fake Job Ads

Thanmai Talluri[1], Haritha Dasari[2]
[1]*M.Tech, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India*
[2]*Professor, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India*

Abstract: The proliferation of fake job advertisements across online recruitment platforms poses a growing challenge to job seekers and undermines trust in digital hiring processes. Existing ensemble learning methods, such as Random Forest, often demonstrate limited capacity to capture complex textual semantics and suffer from high variance and slower convergence. To address these limitations, this work proposes Smart Screen, a robust fake job ad detection system integrating Natural Language Processing (NLP) with the XGBoost algorithm. The system is trained and evaluated on the publicly available Fake Job Postings dataset from Kaggle, combining feature extraction techniques (Bag of Words, TF-IDF, word embeddings) with boosting strategies to enhance accuracy and reduce overfitting. The client-server architecture includes an interface, an ”Analyze Job Post” function for real-time analysis of job post URLs, and an administrative dashboard to monitor and manage user accounts. Experimental evaluation shows notable performance improvements over baseline machine learning models, achieving an accuracy of 97%, F1-score of 96.5%, precision of 96.2%, recall of 96.8%, and an AUC of 98.1%. Additionally, the system includes explainable AI tools such as heatmaps of influential terms and bias-variance analysis reports to improve transparency. This research demonstrates how combining advanced boosting algorithms with practical interface design can strengthen online recruitment security against fraudulent advertisements.

Keywords: XGBoost, NLP, client–server architecture, machine learning, real-time analysis, bias–variance analysis

I. INTRODUCTION

With the widespread adoption of online recruitment platforms, job seekers now enjoy unprecedented access to employment opportunities. However, this convenience has also caused a surge in fraudulent job advertisements, which target especially vulnerable candidates in urgent need of work. Cybercriminals often craft fake postings that closely mimic legitimate vacancies, using persuasive language, false company details, and forged credentials to deceive applicants. As a result, victims can suffer financial loss, identity theft, and increased distrust of digital hiring systems. To address this, machine learning-based fraud detection systems have been widely explored, including ensemble methods such as Random Forest. Although these approaches offer promise, they frequently struggle with the evolving linguistic patterns and complex semantics typical of fake job advertisements. Challenges like high variance, limited generalization to new scams, and insufficient understanding of contextual cues often leave sophisticated fraudulent posts undetected. This dynamic nature of fraud highlights the urgent need for accurate and adaptive detection frameworks capable of keeping up with evolving threats. In response, this work presents Smart Screen, a comprehensive fake job advertisement detection platform that integrates advanced Natural Language Processing (NLP) techniques with the XGBoost algorithm. By extracting rich text features through methods such as Bag of Words, TF-IDF, and word embeddings, Smart Screen captures nuanced patterns characteristic of fraudulent postings. The platform is built around a secure, user-focused architecture that includes a real-time analysis of job post URLs and an administrative interface for monitoring and management. Beyond detection, Smart Screen emphasizes transparency through the integration of explainable AI components, such as visual heat maps highlighting influential terms and bias-variance analysis, helping users and administrators understand why certain posts are flagged. The system is trained and validated on publicly available fake job postings datasets, ensuring that it learns from real-world examples. The key contributions of this research include the design of a robust detection framework combining XGBoost and advanced NLP feature extraction to better differentiate real and fake job postings, the implementation of a secure, user-friendly platform that offers practical, real-time support for job seekers and platform administrators, the integration of interpretable AI tools to improve transparency and foster user trust, and empirical validation demonstrating superior performance compared to traditional machine learning baselines, measured by accuracy, recall, and AUC. By addressing technical limitations in prior systems and focusing on usability and explainability, this study aims to advance the protection of job seekers against online recruitment fraud and enhance confidence in digital hiring environments.

II. RELATED WORK

With the widespread adoption of online recruitment platforms, job seekers now enjoy unprecedented access to employment opportunities [10]. However, this convenience has also caused a surge in fraudulent job advertisements, which target especially vulnerable candidates in urgent need of work [5]. Cybercriminals often craft fake postings that closely mimic legitimate vacancies, using persuasive language, false company details, and forged credentials to deceive applicants [6, 7]. As a result, victims can suffer financial loss, identity theft, and increased distrust of digital hiring systems [10]. To address this, machine learning-based fraud detection systems have been widely explored, including ensemble methods such as Random Forest [1, 4, 11]. Although these approaches offer promise, they frequently struggle with the evolving linguistic patterns and complex semantics typical of fake job advertisements [8]. Challenges like high variance, limited generalization to new scams, and insufficient understanding of contextual cues often leave sophisticated fraudulent posts undetected [9, 8]. The dynamic nature of fraud underscores the urgent need for accurate and adaptive detection frameworks that can keep pace with evolving threats [13].

In response, this work presents Smart Screen, a comprehensive fake job advertisement detection platform that integrates advanced Natural Language Processing (NLP) techniques with the XGBoost algorithm [8, 12]. By extracting rich text features through methods such as Bag of Words, TF-IDF, and word embeddings, Smart Screen captures nuanced patterns characteristic of fraudulent postings [8, 3]. The platform is built around a secure, user-focused architecture that includes a real-time analysis of job post URLs and an administrative interface for monitoring and management [6]. Beyond detection, Smart Screen emphasizes transparency through the integration of explainable AI components, such as visual heat maps highlighting influential terms and bias-variance analysis, helping users and administrators understand why certain posts are flagged [15, 16]. The system is trained and validated on publicly available fake job postings datasets [14], ensuring that it learns from real-world examples.

Prior research on detecting fraudulent job advertisements has primarily explored three methodological directions: rule-based systems, traditional machine learning models, and more recent NLP-driven approaches [1, 2, 5]. Early rule-based methods relied on predefined linguistic patterns and domain heuristics to flag suspicious posts, but these approaches were often rigid and struggled to adapt to evolving fraud tactics [5]. Supervised machine learning algorithms, including Support Vector Machines and Random Forests, improve detection performance by learning discriminative textual and metadata features from labeled datasets, such as the Fake Job Postings dataset [1, 4, 11, 14]. However, these models often faced challenges such as limited semantic understanding and high variance when encountering complex or previously unseen fraudulent content [8, 9]. To address these gaps, recent studies have employed advanced Natural Language Processing techniques, such as TF-IDF and word embeddings, which capture richer contextual patterns and improve detection accuracy [2, 3, 8].

III. METHODOLOGY

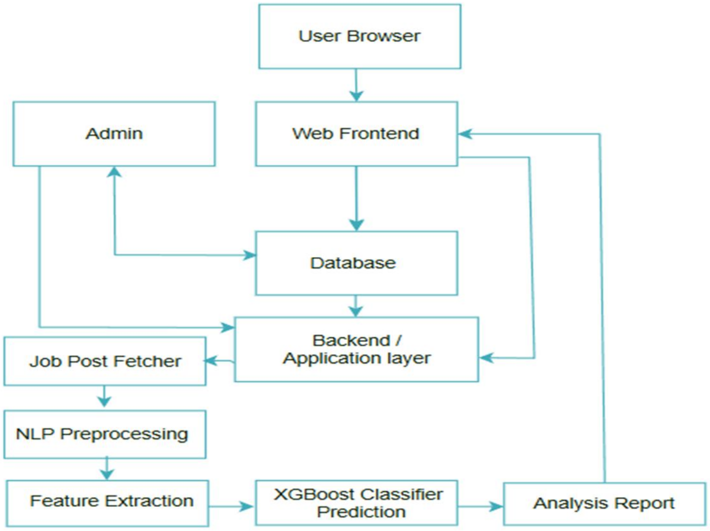


Fig.1: Workflow of Fake Job Ad Detection System

The shown Figure illustrates the complete methodology used in the Fake job ad detection system. The primary objective of this study is to detect fraudulent job advertisements submitted via user-provided URLs. The proposed architecture, illustrated in Fig.1, is designed to process and analyze job postings by leveraging Natural Language Processing (NLP) techniques and an ensemble boosting classifier. This system comprises multiple modular components: a user interface for secure interaction, a backend processing pipeline for data analysis, a database for persistent storage, and an administrative panel for monitoring and management. Together, these components work collaboratively to enable the accurate detection and reporting of fake job ads while ensuring scalability and maintainability. Upon submission of a job post URL by a registered user, the system follows a structured sequence of operations detailed below

1) User Browser: The system interaction commences from the User Browser, through which end users securely access the platform. his entry point facilitates essential operations, including user registration with OTP verification, login authentication, and submission of job post URLs for analysis.
2) Web Frontend: The Web Frontend serves as the primary presentation layer, responsible for rendering dynamic interfaces and managing user interactions. It bridges user requests to the backend services, enabling seamless data exchange. Additionally, it securely channels analysis reports generated by the backend back to the user, ensuring responsiveness and usability.
3) Backend/Application layer: The Backend / Application Layer acts as the computational and logical core of the system. It orchestrates the sequential execution of analytical modules, coordinates data flow, manages user requests, and ensures real-time response generation. This layer operationalizes core algorithms and encapsulates business logic central to fake job ad detection.
4) Job Post Fetcher: Upon receiving a user-submitted URL, the Job Post Fetcher module programmatically retrieves the textual content of the referenced job advertisement. Through techniques such as HTML parsing and content filtering, it isolates relevant descriptive data while excluding extraneous webpage elements, thus producing a clean text corpus for downstream analysis.
5) NLP Preprocessing: The NLP Preprocessing module standardizes and refines the extracted text by applying natural language processing techniques including tokenization, lemmatization, and stopword elimination. This process normalizes linguistic variability, reduces noise, and preserves semantically significant terms, thereby enhancing feature quality.
6) XGBoost Classifier Prediction: At the analytical core of the system, the XGBoost Classifier Prediction module lever ages an ensemble boosting algorithm to model complex non-linear relationships in the feature space. Trained on historical labeled datasets, it predicts the authenticity of the analyzed job post, outputting both the predicted class (genuine or fake) and a confidence score, thus quantifying prediction certainty.
7) Analysis Report: Following classification, the Analysis Report module synthesizes the results into an interpretable and user-friendly report. This document details the classifier's decision, highlights key linguistic features influencing the prediction, and presents a confidence metric, thereby fostering transparency and trust in the system's outputs.
8) Database: The Database module underpins the entire architecture by persistently storing critical data artifacts, including user credentials, analysis histories, feature vectors, and system logs. This repository supports model retraining, auditability, and system performance monitoring, ensuring the platform remains adaptive to evolving patterns in fraudulent advertisements.
9) Admin: The Admin module functions as the system's oversight interface, providing privileged users (administrators) with tools to monitor system health, review user activity, and manage detection models. Its integration with both the Database and Backend / Application Layer enables comprehensive control, ensuring sustained accuracy, compliance, and security. Once the headline is converted into a feature vector, the system extracts contextual metadata from the uploaded image, which may include descriptive captions, tags, or file-level properties that help establish the semantic link between the visual and textual components.

The resulting feature set is then fed into an XGBoost classifier to perform the initial prediction. XGBoost constructs an ensemble of boosted decision trees in a sequen tial manner, minimizing prediction errors through gradient boosting. The predicted probability output of the classifier forms the preliminary label for the news item.

IV. RESULTS

A. Main Page

This page shows the web interface modules, including user registration for new jobseekers and login access for both users and the admin.

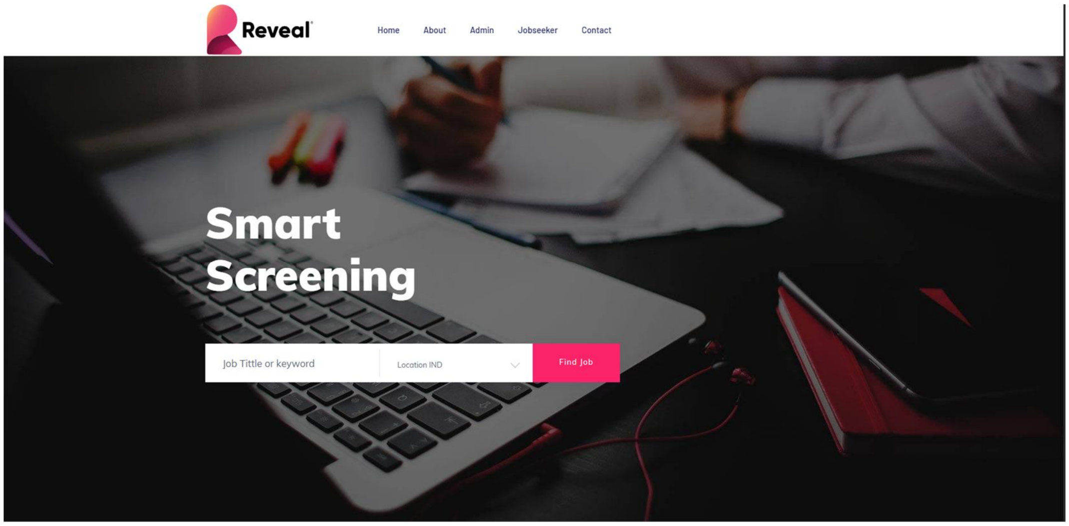


Fig2: Main interface displaying registration and login options for job seekers and the admin panel.

B. Jobseeker Login Page

Upon reaching this page, job seekers enter their registered email and password credentials, which are authenticated against securely stored data in the system’s backend. Successful authentication redirects users to the main dashboard.

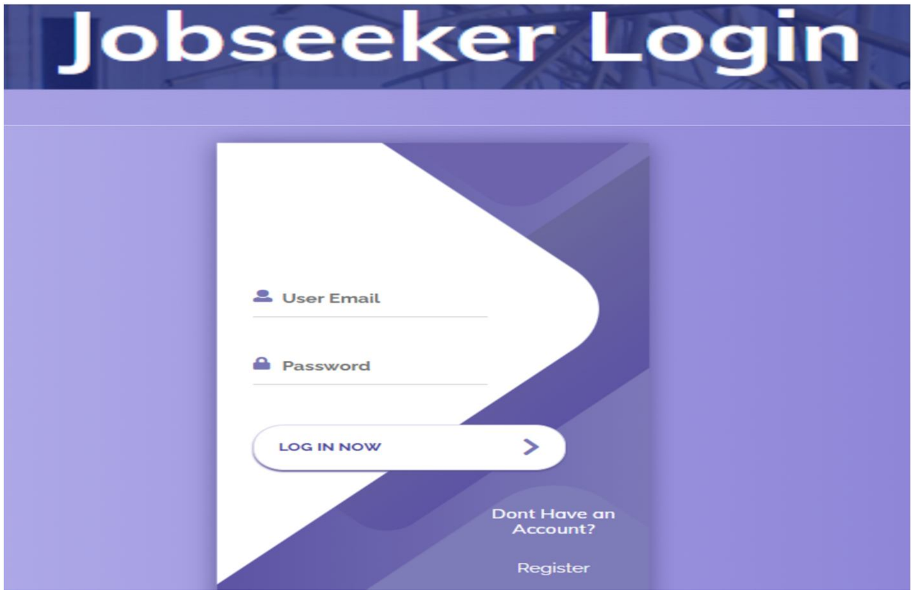


Fig 3: Login interface for jobseekers after successful registration.

C. Admin Login Page

This interface requires authorized credentials to ensure that only verified adminis trators can enter the backend environment. Once authenticated, administrators gain access to advanced features, including viewing user login history, monitoring submit ted job posts, and managing flagged content. The page incorporates security measures such as encrypted password handling and error alerts for incorrect login attempts, thereby safeguarding sensitive administrative functions against unauthorized access. By providing a distinct and secure login page, the system maintains a clear separa tion between regular user activities and administrative controls. This design supports effective oversight, ensures data integrity, and enables proactive intervention against misuse, thus enhancing the reliability and transparency of the fake job advertisement detection platform.

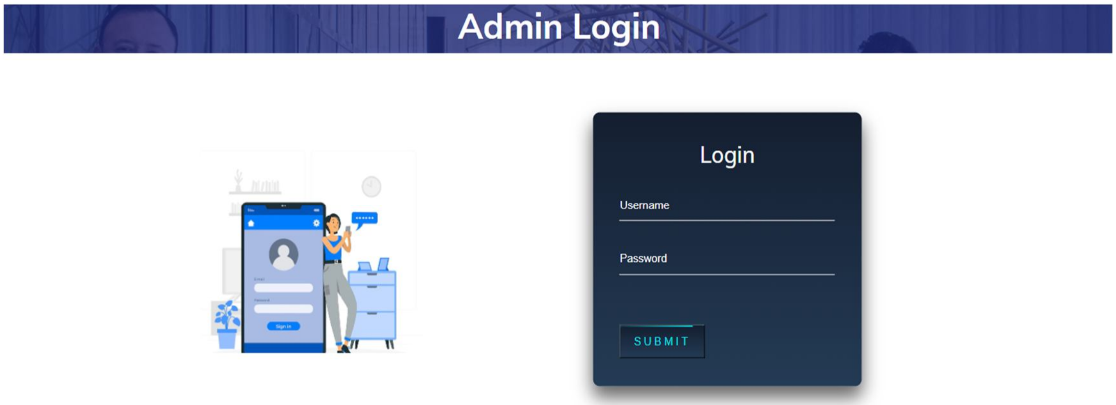
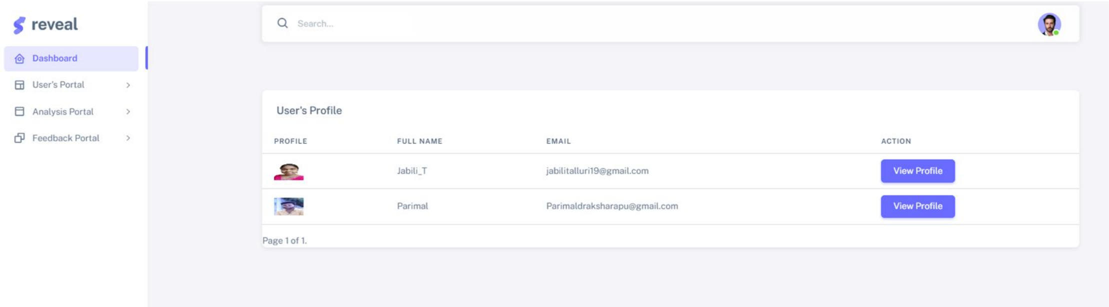


Fig 4: Dedicated login page for administrative users to access the admin dashboard.

D. Profile Page

The Profile Page within the Admin module displays a comprehensive list of registered jobseeker accounts. This feature allows administrators to monitor user activity, review registration details, and manage user profiles as needed. By maintaining visibility into the user base, the admin panel supports continuous oversight and helps ensure responsible use of the fake job ad detection platform. This layer of administrative control contributes to maintaining trust, compliance, and adaptability to evolving fraudulent patterns.



User's Profile

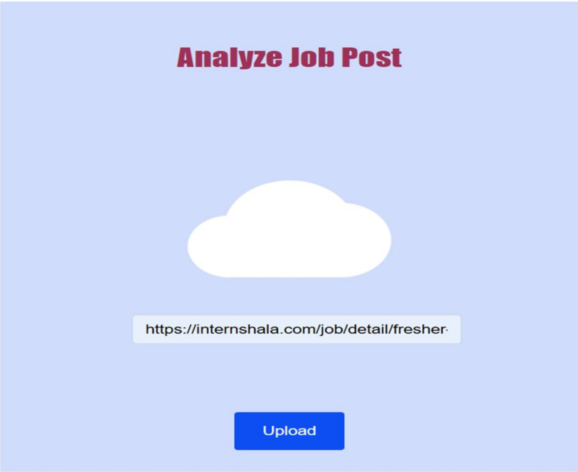
PROFILE	FULL NAME	EMAIL	ACTION
	Jabili_T	jabilitalluri19@gmail.com	View Profile
	Parimal	Parimaldraksharapu@gmail.com	View Profile

Page 1 of 1.

Figure 5: Admin view displaying user profiles registered in the system.

E. Analyze Job Post Page

The Upload Job Post URL feature enhances usability by allowing users to quickly ana lyze online job advertisements without manually copying content. Instead of uploading a document or typing details, job seekers can simply submit the direct URL of the job posting they wish to verify. The system automatically retrieves the content from the given link.

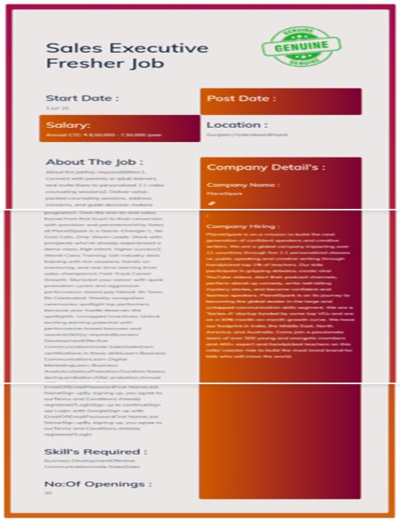


https://internshala.com/job/detail/fresher-

Fig5: Admin view displaying user profiles registered in the system.

F. Analysis Report- Genuine

Upon successful analysis, a submitted job post—whether uploaded as text, file, or URL—the system generates an Analysis Report that presents the prediction results in an intuitive format. When a job post is classified as genuine, the report displays key details such as the job title, company name, posting date, and extracted job description to provide full context. Additionally, the report prominently features a "Genuine" verification stamp or badge, visually confirming the authenticity of the post. This design helps job seekers quickly understand the decision and builds trust by offering both a detailed textual breakdown and a clear visual indicatorsignificant seismic events. The model's strength in correctly identifying high-magnitude earthquakes has substantial implications for public safety, early warning systems, and disaster preparedness. Future research can further enhance this model by incorporating real-time geospatial data, seismic wave analysis, and advanced deep learning techniques to improve early detection rates and reduce false alarms.



Sales Executive Fresher Job

GENUINE

Start Date :
3 Jul' 25

Post Date :

Salary:
Annual CTC: ₹ 6,50,000 - 7,50,000 /year

Location :
Gurgaon,Hyderabad,Bhopal

About The Job :
About the jobKey responsibilities:1. Connect with parents or adult learners and invite them to personalized 1:1 video counseling sessions2. Deliver value-packed counseling sessions, address concerns, and guide decision makers programs3. Own the end-to-end sales funnel from first touch to final conversion with precision and persistenceWhy Sales at PlanetSpark is a Game-Changer:1. No Cold Calls, Only Warm Leads: Work with prospects who've already experienced a demo class, high intent, higher success2. World-Class Training: Get industry-best training with live sessions, hands-on mentoring, and real-time learning from sales champions3. Fast-Track Career Growth: Skyrocket your career with quick promotion cycles and aggressive performance-based pay hikes4. Be Seen, Be Celebrated: Weekly recognition ceremonies spotlight top performers because your hustle deserves the spotlight5. Uncapped Incentives: Unlock exciting earning potential with performance-based bonuses and rewardsSkill(s) requiredBusiness DevelopmentEffective CommunicationInside SalesSalesEarn certifications in these skillsLearn Business CommunicationLearn Digital MarketingLearn Business AnalyticsSalaryProbation:Duration:Salary during probation:After probation:Annual [illegible]
EmailOREmailPasswordFirst NameLast NameSign upBy signing up, you agree to ourTerms and Conditions.Already registered?LoginSign up to continueSign up/ Login with GoogleSign up with EmailOREmailPasswordFirst NameLast NameSign upBy signing up, you agree to ourTerms and Conditions.Already registered?Login

Skill's Required :
Business DevelopmentEffective CommunicationInside SalesSales

No:Of Openings :
30

Company Detail's :

Company Name :
PlanetSpark

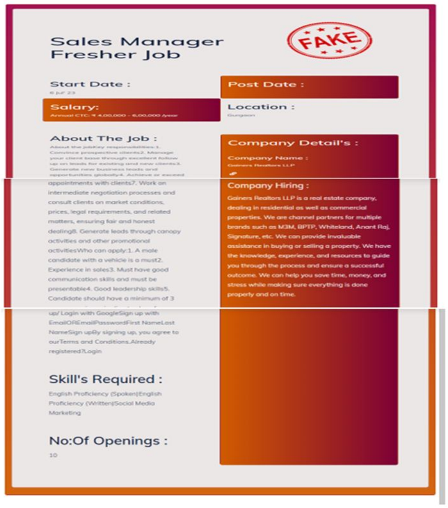
:

Company Hiring :
PlanetSpark is on a mission to build the next generation of confident speakers and creative writers. We are a global company impacting over 11 countries through live 1:1 personalized classes on public speaking and creative writing through handpicked top 1% of teachers. Our kids participate in gripping debates, create viral YouTube videos, start their podcast channels, perform stand-up comedy, write nail-biting mystery stories, and become confident and fearless speakers. PlanetSpark is on its journey to becoming the global leader in the large and untapped communication skills segment. We are a "Series-A" startup funded by some top VCs and are on a 30% month-on-month growth curve. We have our footprint in India, the Middle East, North America, and Australia. Come join a passionate team of over 500 young and energetic members and 400+ expert and handpicked teachers on this roller coaster ride to build the most loved brand for kids who will move the world.

Fig6: Result Report of the job post is Genuine

G. Analysis Report- Fake

When the system identifies a job advertisement as fake, the Analysis Report presents this finding prominently, along with a confidence score reflecting the model's cer tainty in its classification. To enhance transparency, the report also enumerates the most influential keywords and metadata features that triggered the fake classification, such as suspicious phrases promising unrealistic benefits or patterns characteristic of known scam attempts. By making the rationale behind its decision explicit, the system empowers users to understand the underlying risks, promotes informed decision making, and raises awareness about the linguistic and structural signals often exploited by fraudulent postings.



Sales Manager Fresher Job

FAKE

Start Date :
6 Jul' 23

Post Date :

Salary:
Annual CTC: ₹ 4,00,000 - 6,00,000 /year

Location :
Gurgaon

About The Job :
About the jobKey responsibilities:1. Convince prospective clients2. Manage your client base through excellent follow up on leads for existing and new clients3. Generate new business leads and opportunities globally4. Achieve or exceed
appointments with clients7. Work on intermediate negotiation processes and consult clients on market conditions, prices, legal requirements, and related matters, ensuring fair and honest dealing8. Generate leads through canopy activities and other promotional activitiesWho can apply:1. A male candidate with a vehicle is a must2. Experience in sales3. Must have good communication skills and must be presentable4. Good leadership skills5. Candidate should have a minimum of 3
up/ Login with GoogleSign up with EmailOREmailPasswordFirst NameLast NameSign upBy signing up, you agree to ourTerms and Conditions.Already registered?Login

Skill's Required :
English Proficiency (Spoken)English Proficiency (Written)Social Media Marketing

No:Of Openings :
10

Company Detail's :

Company Name :
Gainers Realtors LLP

Company Hiring :
Gainers Realtors LLP is a real estate company, dealing in residential as well as commercial properties. We are channel partners for multiple brands such as M3M, BPTP, Whiteland, Anant Raj, Signature, etc. We can provide invaluable assistance in buying or selling a property. We have the knowledge, experience, and resources to guide you through the process and ensure a successful outcome. We can help you save time, money, and stress while making sure everything is done properly and on time.

Fig7: Result Report of the job post is Fake

V. CONCLUSION

A distinct contribution of this work is the development of a web-based interface, which transforms the technical solution into a practical tool usable by job seekers in real time. This direct accessibility ensures that individuals without technical expertise can quickly assess the authenticity of job postings, thereby analysis reducing the risk of fraud and enhancing overall trust in digital hiring platforms. Our experimental analysis demonstrates that the integration of these methods results in a robust and adaptable detection system, capable of handling diverse text patterns commonly seen in deceptive job ads. The model's effectiveness becomes espe cially valuable in contexts where fraudulent postings use persuasive or ambiguous language, posing challenges to standard detection approaches. Nonetheless, the study acknowledges limitations, such as potential sensitivity to dataset language and domain shifts, which could affect performance in rapidly evolving fraud scenarios.

Future research directions include extending the system to support multiple languages, employing deep semantic models to better capture contextual nuances, and implementing dynamic retraining strategies to adapt to new fraud tactics. In general, this research illustrates that a combined approach- taking advantage of the strengths of NLP and ensemble learning within an accessible web interface can significantly advance the detection of fake job advertisements. This work offers both theoretical contributions and tangible benefits, ultimately helping to protect job seekers from deceptive recruitment practices in an increasingly digital world.

REFERENCES

[1] Kumar A, Garg N (2020) Detecting fraudulent job postings using machine learning. Procedia Comput Sci 167:2101–2110
[2] Malhotra P, Arora A (2021) Fake job post detection using natural language processing and ensemble learning. Int J Adv Comput Sci Appl 12(6):45–52
[3] Zhang Y, Zheng X (2021) Detection of fake job advertisements using BERT and deep neural networks. Neural Comput Appl 33:15565–15578
[4] Gupta R, Sharma V (2020) Ensemble methods for identifying fraudulent online job postings. J Comput Sci Technol 35(5):1021–1033
[5] Wang H, Liu J (2019) Using text mining and classification techniques to detect recruitment fraud. Expert Syst Appl 125:205–215
[6] Chen L, Zhao Y (2022) Automated detection of deceptive job ads on social networks. Soc Netw Anal Min 12(1):25
[7] Patel M, Kumar S (2021) Hybrid machine learning framework for fake job offer detection. Comput Electr Eng 89:106898
[8] Sharma N, Verma P (2020) Application of NLP and XGBoost for detecting fake recruitment advertisements. Int J Inf Secur 19(6):567–579
[9] Li F, Wu Q (2021) A comparative study on machine learning algorithms for detecting fake job posts. IEEE Access 9:115678–115689
[10] Alghamdi B, Alharby F (2019) An Intelligent Model for Online Recruitment Fraud Detection. J Inf Secur 10(03):155–176
[11] Breiman L (2001) Random Forests. Mach Learn 45(1):5–32
[12] Natekin A, Knoll A (2013) Gradient Boosting Machines: A Tutorial. Front Neurorobot 7:21
[13] Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media. ACM SIGKDD Explor Newsl 19(1):22–36
[14] Bansal S (2020) [Real or Fake] Fake Job Posting Prediction [Dataset]. Kaggle
[15] Vieira SM, Kaymak U, Sousa JMC (2010) Cohen's Kappa Coefficient as a Per formance Measure for Feature Selection. In: 2010 IEEE World Congress on Computational Intelligence. IEEE, pp 1–7
[16] Biggio B, Corona I, Fumera G, Giacinto G, Roli F (2011) Bagging classifiers for f ighting poisoning attacks in adversarial classification tasks. Lect Notes Comput Sci 6713:350–359