**Job Sniffer: AI-Powered Fake Job Post Detector**

**GBA6410 – Text Mining and Social Media**

**Final Project Report**

Group 6

Peiching Tsai

Tingwei Koo

Jacob Rudder

Cedric Wong

Stephen Deng

July 22, 2025

**Introduction**

With the growing digitalization of the job market, online recruitment platforms have become essential tools for connecting job seekers with employers. However, this increased reliance has also opened the door to a surge in fake job postings designed to exploit vulnerable applicants. These fraudulent listings often closely mimic legitimate ones, making them difficult to detect through traditional means.

This project aims to develop an automated system that identifies such scams by analyzing the linguistic and structural patterns of job descriptions. Our central research question is: What textual and structural features can effectively distinguish fake job postings from legitimate ones using text mining and machine learning techniques?

Addressing this issue is critical, as fake job listings pose serious risks, including identity theft, financial fraud, and time loss for job seekers. Despite the scale of the problem, current detection efforts rely heavily on manual reporting, which is inefficient and reactive. To address this gap, we propose a scalable and AI-driven solution, *Job-Sniffer*, which automates scam detection by leveraging linguistic cues embedded in job descriptions. This research contributes to ongoing efforts in cybersecurity, natural language processing (NLP), and digital trust by offering a practical and data-driven approach to safeguarding online job seekers.

**Background Research**

1. Why do fake job postings exist?

Fake job postings exist due to a range of deceptive motivations, many of which are rooted in exploitation, manipulation, or artificial engagement. These postings are designed to appear legitimate while pursuing hidden agendas such as phishing for personal data, harvesting resumes, scamming users with fraudulent offers, or inflating traffic metrics on job boards and platforms.

A primary driver behind fake postings is phishing and identity theft. According to the Better Business Bureau (2021), millions of job seekers across the U.S. and Canada are targeted by fraudulent listings each year, leading to identity theft, fake check scams, and even illegal job traps. The FBI’s IC3 reported a steady rise in monetary losses from these scams, indicating the growing sophistication and prevalence of such schemes.

Another motivation is data harvesting. Many fake employers are more interested in gathering sensitive applicant data, such as names, phone numbers, emails, and employment history, than actually hiring. This information can then be sold or exploited in further scams. Some postings also exist to manipulate brand perception, portraying a company as growing or hiring to create a façade of expansion, even if there are no real openings.

Furthermore, companies or individuals may post fake listings to inflate job board engagement or test applicant reactions to proposed roles, salaries, or titles without the intent of filling any positions. In the research presented by Vrinda et al. (2024), such listings often include linguistic red flags like urgency-driven language (“apply now”), promises of unrealistic compensation, or vague job descriptions, clues that can be detected with machine learning.

Technologically, fake postings exploit the lack of robust verification on many platforms. As Pillai (2023) and Sahana & Ashwini (2024) demonstrate, many of these fraudulent listings rely on persuasive yet generic language, often reusing the same structure across multiple fake jobs. These characteristics make them ripe for detection using natural language processing (NLP) and machine learning (ML), which can uncover subtle yet consistent patterns in text.

Finally, from a broader social lens, the high demand for employment, particularly during economic downturns, creates a ripe environment for scammers. With the average job search taking over 240 days (Pathrise, 2024), job seekers, especially those desperate for work, are more vulnerable to deceptive offers, increasing the success rate of such scams.

Understanding these motivations is crucial for developing effective detection methods. It not only provides necessary context for identifying linguistic cues in fraudulent listings, but also reinforces the urgency of protecting job seekers and digital hiring ecosystems from harm. By uncovering the “why,” we can better inform the “how” in combating job post fraud through intelligent, ethical technology.

1. What Linguistic Features Distinguish Fake Job Postings?

Given the natural state of fake job posting, this study investigates the linguistic and structural characteristics that distinguish fake job postings from legitimate ones, leveraging text mining techniques for automated detection. We analyze prior research on deceptive job ads, identify key patterns of vagueness and inconsistencies between job descriptions, and evaluate the effectiveness of the classification model.

Some key linguistic characteristics noted from both prior work and our study of fake job posting included the use of “buzz words”, such as “work from instantly” and “easy money”, and vague descriptions of job that does not contain information of the specific role and responsibilities of the position. In addition, fake posting is also characterized by metadata anomalies such as the usage of unprofessional contact information and urgent tactics. These unique characteristics of fake job postings allow them to be identified with NPL and ML.

A study conducted by University of Illinois completed research on fake job posting using Count Vectorization, an NLP technique used to convert a collection of text documents into a matrix of token counts to feed their model input. This study will take a similar approach. The preprocessing step of the study aims to identify the most common bi-grams used in real job postings and differentiate them from those of fake postings. Similar to previous findings, we noted that fake job postings including high occurrences of lackluster and low requirement terms including “basic knowledge” and “flexible hours”. In contrast, real posting contains terms of more specific requirements such as “experience” and “communication skills”. The difference in the concentration of these key terms would be visualized in a vectorization process such as TF-IDF, which will serve as the basis for the model to identify fake and real job postings.

**Source Data and Preprocessing**

This study utilizes two publicly available datasets from Kaggle that contain labeled job postings. The first dataset, compiled by Sri Sai Suhas Sanisetty, consists of approximately 10,000 records labeled as fraudulent. It includes multiple textual fields such as title, company\_profile, and description, which are essential for text-based classification. The second dataset, contributed by Hassan Legends, comprises 17,880 job postings—17,014 labeled as legitimate and 866 as fake—also with structured textual features. These two datasets were selected due to their relevance and rare inclusion of binary fraud labels, making them particularly suitable for supervised text mining.

To improve representativeness and data diversity, the two datasets were merged, producing a combined corpus of 27,880 postings. The class distribution is approximately 61% real and 39% fake, indicating moderate imbalance. This imbalance was mitigated in later modeling through oversampling.

Prior to analysis, missing values were examined. Records lacking entries in the description field—critical for downstream NLP—were removed. For other fields such as company\_profile or benefits, missing values were replaced with empty strings to retain data volume and prevent imputation bias. After cleaning, the final dataset consisted of 27,879 records with ten key variables: title, location, salary\_range, company\_profile, description, requirements, benefits, employment\_type, industry, and fraudulent.

Text preprocessing was applied to four primary fields: company\_profile, description, requirements, and benefits. The following steps were performed:

(1) lowercasing for consistency,

(2) removal of non-alphabetic characters and extra whitespace,

(3) tokenization into word-level units,

(4) stopword removal using NLTK’s English stopword list,

(5) lemmatization to normalize word forms.

These preprocessing steps produced a clean, semantically consistent dataset suitable for feature extraction and classification.

**Descriptive Results of Data**

|  |  |  |
| --- | --- | --- |
| Column | Real Top-2-grams (count) | Fake Top-2-grams (count) |
| company\_profile | full time (1412)  business process (1291)  around world (1227)  increase productivity (1159)  document communication (1148)  high quality (984)  long term (801)  customer satisfaction (797)  new york (765)  amp secure (750) | plc established (596)  inc established (560)  llc established (560)  son established (553)  ltd established (539)  group established (524)  smith established (136)  johnson established (126)  signing bonus (112)  aptitude staffing (107) |
| description | customer service(2759)  full time (1664)  social medium (1492)  team member (1278)  ideal candidate (1123)  join team (1077)  fast paced (1025)  communication skill (1012)  day day (855)  work closely (849) | earn week (10002)  week immediate (10000)  immediate hiring (10000)  hiring contact (10000)  hotmail com (1718)  yahoo com (1669)  gmail com (1641)  customer service (181)  oil gas (147)  full time (145) |
| requirements | year experience (3255)  communication skill (3052)  customer service (1766)  ability work (1338)  minimum year (1232)  fast paced (1207)  skill ability (1200)  experience working (1169)  bachelor degree (1106)  computer science (1095) | basic knowledge (10004)  degree required (10003)  flexible hour (10001)  required flexible (10000)  knowledge degree (1094)  communication skill (116)  year experience (112)  ability work (104)  skill ability (86)  high school (83) |
| benefits | competitive salary (1197)  benefit package (955)  offer competitive (866)  dental vision (858)  full time (841)  job description (774)  medical dental (758)  see job (728)  health dental (505)  paid time (431) | sign bonus (2068)  free travel (1994)  free meal (1986)  remote work (1986)  work opportunity (1986)  flexible hour (1973)  get started (93)  c b (78)  benefit package (69)  work life (69) |

In analyzing the company profile section, we observed a stark contrast between real and fake job postings. Real postings tend to present a multidimensional view of the organization, emphasizing core values such as *customer satisfaction*, *increased productivity*, and *high quality*. They also outline the operational scope with phrases like *around the world*, *business process*, and *long term*, as well as specifying work modality and location, using terms such as *full time*, *New York*, and *amp secure*. In contrast, fake postings often rely on boilerplate historical phrases, with over 80% of the top bi-grams structured as “[CompanySuffix] established” (e.g., *LLC established*, *PLC established*), yet they fail to provide actual founding dates or mission statements. Additionally, scattered tokens such as *signing bonus* and *aptitude staffing* further indicate a templated and generic approach to content generation.

The job description sections show similar divergence. Real job postings emphasize clarity in roles and responsibilities, as well as the importance of teamwork. Phrases such as *join team*, *team member*, and *ideal candidate* highlight collaborative environments, while descriptions like *fast paced*, *communication skill*, and *work closely* provide insight into daily expectations. Some postings even include truncated forms like *day day*, derived from *day-to-day*. In contrast, fake postings often create a sense of urgency or financial temptation, using terms like *immediate hiring*, *week immediate*, and *earn weekly*. These listings frequently include personal email domains such as *hotmail.com*, *gmail.com*, or *yahoo.com*, suggesting a lack of corporate infrastructure. The presence of incongruous terms like *oil gas* further implies indiscriminate copy-pasting from unrelated content.

In the requirements section, real job postings tend to be more specific and measurable. They include clear experience thresholds, using terms such as *year experience*, *minimum year*, and *experience working*. Educational and skill-based qualifications are also common, including phrases like *bachelor degree*, *computer science*, and *communication skill*. Contextual terms such as *fast paced* also appear frequently, indicating the working environment. Conversely, fake job postings rely heavily on vague and catch-all phrases such as *basic knowledge*, *degree required*, and *knowledge degree*. The criteria are often overly flexible, with terms like *flexible hour*, *required flexible*, and *high school* designed to appeal to a broad, less qualified applicant pool. Even legitimate qualifiers like *communication skill* appear far less frequently and are often overshadowed by more generic language.

Lastly, in the benefits section, real job postings tend to list concrete and industry-standard perks. They offer clarity regarding compensation with phrases such as *competitive salary*, *benefit package*, and *offer competitive*. Health-related benefits are commonly noted, with frequent mentions of *medical dental*, *dental vision*, and *health dental*. Work-related benefits such as *full time* and *paid time* are also emphasized. In contrast, fake job postings often use overly attractive yet vague incentives to lure applicants. These include promises of *sign bonus*, *free travel*, *free meal*, as well as lifestyle perks like *remote work*, *flexible hour*, and *work opportunity*. Although the term *benefit package* is sometimes used in fake postings, the surrounding context often lacks the depth and specificity found in real job listings.

**Text‐Mining Methodology**

**Rationale**

The proliferation of fraudulent job postings across online recruitment platforms has created an urgent need for scalable, intelligent detection systems. While early research efforts have employed traditional machine learning techniques—such as TF-IDF or Bag-of-Words (BoW) representations paired with classifiers like Decision Trees or Support Vector Machines (e.g., Hanif et al., 2024; Boka, 2024)— these approaches just look at the words on the surface, so they struggle when scammers get sneaky or try to hide what they’re really saying. Although computationally efficient, they struggle to adapt to the rapidly evolving linguistic strategies used in job scams.

To address these limitations, more recent studies have turned to deep learning architectures such as Bi-LSTM (Pillai, 2023) and transformer-based models like fine-tuned DistilBERT and Fraud-BERT (Sree Narayana College, 2024; Fraud-BERT, 2025). These models offer deeper semantic understanding and have demonstrated state-of-the-art performance in text classification. However, their black-box nature, high resource requirements, and lack of interpretability remain significant barriers to adoption in real-world pre-screening systems.

In response, this study introduces *Job-Sniffer*, a hybrid fraud detection framework that bridges the gap between interpretability and semantic depth. Our system integrates three key components:

(1) rule-based detection using red-flag n-grams and email domain heuristics,

(2) contextual embeddings via Sentence-BERT to capture nuanced semantics, and

(3) unsupervised anomaly detection using Isolation Forest to identify outliers beyond known scam patterns.

This layered architecture not only improves detection accuracy, but also enhances transparency—facilitating human-in-the-loop verification and enabling deployment in trust-critical environments. By balancing interpretability, precision, and scalability, *Job-Sniffer* advances the field beyond prior lexical or end-to-end approaches and delivers a more adaptable, real-world solution for combating job post fraud.

**Methodology Overview**

*Job-Sniffer* employs a stacked ensemble architecture comprising rule-based red flag detection, semantic modeling, and anomaly detection. This hybrid framework enhances robustness against both known and novel fraud patterns while maintaining transparency through interpretable intermediate outputs that support human-in-the-loop validation.

Following standard text preprocessing, the system constructs a rich feature set by combining surface-level patterns with deeper semantic cues. High-frequency n-grams commonly found in scam postings (e.g., “LLC established,” “immediate hiring”) are extracted to form a red-flag dictionary, which is used to generate binary indicators for suspicious textual elements. Concurrently, TF-IDF embeddings are computed to capture term-level relevance across postings, while Sentence-BERT generates sentence-level contextual embeddings that model deeper semantic relationships.

The rule-based detection layer identifies explicit red flags, such as known scam phrases or the presence of personal email domains like “gmail.com.” Additional document-level features, including lexical diversity—approximated via Shannon entropy—and readability scores, contribute further structure for interpretability and detection.

An unsupervised anomaly detection layer, trained exclusively on verified job postings using Isolation Forest, captures statistical outliers in the feature space that may correspond to novel or obfuscated fraud patterns. The final decision layer consists of a Random Forest classifier trained on labeled examples, using the concatenated outputs of all prior layers. At inference time, the system aggregates the outputs using a weighted ensemble: 30 percent from rule-based flag scores, 30 percent from anomaly scores, and 40 percent from supervised classification probabilities.

This layered architecture offers a scalable and interpretable solution capable of detecting both conventional and zero-day fraudulent job postings in dynamic recruitment environments.

**Implementation and Initial Results**

To evaluate the effectiveness of the proposed *Job-Sniffer* agent, we conducted experiments on a labeled dataset of job postings. The data was stratified and split into training and test subsets at a 70:30 ratio, preserving the class distribution. The model was trained on the training set using a combination of rule-based features, TF-IDF representations, contextual embeddings from Sentence-BERT, and outlier scores from Isolation Forest. Red-flag n-grams were dynamically extracted from the fraudulent subset to support binary heuristic flagging during both training and inference.

After training, the agent was applied to the test set, where each posting was scored using the final ensemble method. A threshold of 0.5 was used to convert the final fraud probability score into binary predictions. The performance was evaluated using standard classification metrics, including ROC-AUC, precision, recall, and F1-score.

The results indicate that the *Job-Sniffer* agent achieved strong predictive performance, with a ROC-AUC score of 0.994, suggesting excellent separability between real and fraudulent postings. The model reached perfect precision (1.000), indicating no false positives under the given threshold, and a recall of 0.920, meaning the majority of actual fraudulent listings were successfully identified. The overall F1-score was 0.958, reflecting a well-balanced trade-off between precision and recall.

The ROC curve, shown in Figure 1, illustrates a sharp rise toward the top-left corner, confirming the model’s high discriminative ability. Additionally, the confusion matrix, shown in Table 1, demonstrated that the classifier produced very few false negatives and no false positives in this configuration, highlighting the system’s reliability in high-precision use cases such as pre-screening for manual fraud investigation.

A graph of a graph with a line

AI-generated content may be incorrect.

Figure 1: The ROC curve of the proposed agent.

|  |  |  |
| --- | --- | --- |
|  | Predicted Real (0) | Predicted Fake (1) |
| Actual Real (0) | 5104 | 0 |
| Actual Fake (1) | 261 | 2999 |

Table 1: The confusion matrix based on a 0.5 decision threshold.

To address potential concerns regarding class imbalance and overfitting—especially considering the model’s high ROC-AUC score—we ensured stratified sampling was applied during the train-test split to preserve class distributions. Additionally, the original datasets had an imbalance (approximately 61% real vs. 39% fake), so we applied oversampling techniques during training to mitigate bias.

While the ROC curve rises sharply, indicating strong separability, this may reflect high sensitivity to known red flags. To ensure generalizability, further validation using k-fold cross-validation and out-of-sample data is recommended. Moreover, no single feature dominated the ensemble; the inclusion of anomaly detection (Isolation Forest) and rule-based flags was designed to reduce overfitting and encourage model interpretability.

These initial results validate the robustness of the hybrid ensemble framework and support the viability of *Job-Sniffer* as a scalable and interpretable tool for automated fraud detection in online recruitment platforms.

**Discussion, Limitation and Future Research**

Our *Job-Sniffer* Agent was able to obtain a 92% or greater on several commonly used performance metrics such as precision, recall and ROC-AUC. This indicates that the models are accurate and useful for implementation in job board websites as originally discussed. Sites such as Linkedin can use our multi-layer agent or a similar setup to great effectiveness in mass reporting fraudulent job postings without the need for user-submitted reports.

Due to the difference in size and collection method between the two original datasets which were merged, the results may be biased in favor of the larger dataset. Our model does perform well on existing data, but further testing may be needed for live usage. Since the data is from job postings instead of job applicants, user privacy was somewhat less of a concern as there was no personal data present in the dataset. We also protected user privacy by removing potentially sensitive columns such as location data and listed salaries outside of the text-based descriptions.

We were somewhat limited in this project by our datasets, neither of which individually had what we needed. It took merging two different datsets – one with fake job postings, and another with mostly real ones, to have both the volume and diversity of data necessary for analysis. A future project would ideally work with big data on the scale of millions of job applications rather than tens of thousands and potentially do so in a real-time environment where the agent is constantly being updated. The number and complexity of the models could also be further increased or refined, modifying the layers of the agent’s methodology to update it as new methodologies and models are created.

**Conclusion**

This study set out to identify which textual and structural features are most effective for distinguishing fake job postings from legitimate ones. By developing and evaluating the *Job-Sniffer* agent on labeled datasets, we demonstrated that combining interpretable linguistic heuristics with semantic modeling and anomaly detection yields robust and scalable performance. The final system achieved a ROC-AUC of 0.994, perfect precision, and an F1-score of 0.958, confirming its effectiveness in detecting deceptive listings.

Our analysis further revealed that job postings exhibiting vague role descriptions, excessive urgency cues (e.g., “immediate hiring”), unrealistic incentives (“earn weekly”, “free travel”), and the use of personal email domains (e.g., gmail.com, yahoo.com) are particularly indicative of fraudulent behavior. Additionally, structural signals such as low lexical diversity, minimal qualification requirements, and generic company profiles lacking concrete organizational context were found to be strong predictors of scam listings.

Beyond its predictive performance, the proposed framework contributes to the broader goal of building trustworthy online recruitment platforms by offering a detection system that is both interpretable and adaptable. Future work may extend this research to multilingual contexts, adversarial robustness, and integration with platform-level screening tools for large-scale deployment. Beyond its predictive accuracy, the system contributes to ongoing efforts in ethical AI by offering both explainability and adaptability. These findings affirm the feasibility of using text mining and machine learning to automate fraud detection at scale. Future work may explore adversarial robustness, multilingual generalization, and cross-platform adaptation to further extend the system’s utility in real-world hiring environments.

**Reference**

Better Business Bureau. (2021, September). *Job scams: BBB study finds job scams increased during pandemic and warns job seekers to verify employment offers to avoid illegal jobs, identity theft and fake checks*. BBB International Investigations Initiative. <https://www.bbb.org/>

Boka, M. (2024, April). *Predicting Fake Job Posts Using Machine Learning Models*. *SSRN*.

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4807145>

*Fraud‑BERT*: Transformer‑based contextual detection of fraudulent job posts. (2025).

*Journal of Information Retrieval*. <https://jespublication.com/upload/2023-V14I209.pdf>?

Mohd Hanif, A. H., Maarop, N., Kamaruddin, N., & Samy, G. N. (2024). Machine learning

approach in predicting fraudulent job advertisement. *International Journal of Academic Research in Business and Social Sciences*, *14*(1). <https://doi.org/10.6007/ijarbss/v14-i1/20532>

Pathrise. (n.d.). *How long does it take to find a job?* Pathrise. Retrieved June 23, 2025, from <https://www.pathrise.com/guides/how-long-does-it-take-to-find-a-job/>

Pillai, Aravind Sasidharan. (2023). Detecting Fake Job Postings Using Bidirectional LSTM.

<https://www.researchgate.net/publication/369823698_Detecting_Fake_Job_Postings_Using_Bidirectional_LSTM>

Ranparia, *et al.* (2024). *GloVe + RNN applied to LinkedIn data for scam detection*

[https://www.irjmets.com/uploadedfiles/paper//issue\_8\_august\_2024/61269/final/fin\_irjmets1725172896.pdf](https://www.irjmets.com/uploadedfiles/paper/issue_8_august_2024/61269/final/fin_irjmets1725172896.pdf)?

Sree Narayana College. (2024). *Fine‑tuned DistilBERT for job‑scam detection*

<https://www.ijirset.com/upload/2024/september/113_Fraudulent.pdf>