Fake Job Detection

[1]Amrutha P, MCA student,

PES Institute of Technology and Management, Shivamogga, Karnataka, India.

[2]Ms. Kavya H V, Assistant Professor, MCA,

PES Institute of Technology and Management, Shivamogga, Karnataka, India

Abstract

During the pandemic there is increase in the number of jobs advertised online across different job portals So, deceptive job posting prediction task will be big. The growth of online job boards has clearly made it is more accessible to potential employees and at simultaneously it has opened up the opportunity to fraudsters to commit recruitment scams by listing fake advertisements. This type of fake posting does not only waste time of the applicant but also endangers personal and financial information. As a counter to that, the current project will present a Fake Job Requirement Prediction System is developed by machine learning and natural language processing that can analyse advertisements and categorize them as valid or fake. Its design is implemented on the MERN stack which provides the dynamic interface towards user and Fast API was embedded into its backend which provides faster data processing. The system consists of two primary modules, User and Admin. More than one URLs of job posting are submitted through the interface on running this, the system checks the threat found in the link via Google Safe Browsing. In the case that the link is proved to be safe, web scraping is used where the information is extracted depending on the role in this case, the information is company name, position, salary, job type and other related information. The aid of the Gemini API is also useful in the retrieved features since it provides other indicator variables relevant to the legitimacy of the firm. This consolidated document is used in a machine learning file that identifies authenticity of the employment. The admin module is helpful to the administrator in tracking the work of the users and everything detected. The protection of the job seeker against online recruitment cheat, as well as the betterment of a more secure job market through optimization of real-time web technologies coupled with the supplementary designation models, is the aspiration that the system aspires to attain.

Keywords

Genuine, theft, fake, personal information, precisely detected, Machine Learning, NLP Classification.

1. Introduction

The emergence of digital requisition services has made the method of obtaining staff much faster both by the employees and the applicants but this same change in technology also led to a massive increase in fake job advertisements that are designed to defraud the job seeker and get hold of his or her highly confidential information. Manual inspection to discover such fake advertisements is ineffective and impracticable as the amount of posting through web portals, social media sites and email communication is unprecedented. This paper will present a Fake Job Requirement Prediction System that will automate the process of identifying and categorizing employment opportunities that are either legitimate or fake. The system uses the hybrid model which implies the combination of by machine learning and natural language processing coupled with the modern web development stacks to be able to provide real-time study of the job post. MERN stack will handle user-interface and user-interaction elements and Fast API will present the backend infrastructure. In the event that an employment URL is provided, Google Safe Browsing API is applied in analysing the security of the link. Supposing the URL of interest is secure, the concerned data is received using web scraping the corresponding data are job title, salary range, the name of the company, and the type of listing. The information obtained goes to the Gemini API to verify whether the posting is real or not. The system consists of two parts. The User module allows the job applicants to feed job URLs to investigate the same and the admin module allows the system administrators to manage the user accounts along with to track the effect of detection. Collectively, these services are to form an intelligent, programmatic security measure and they are designed to secure users against scam job listings on the internet and in the process, ensure a secure and trusted online working environment is obtained.

2. Literature Survey

Clark, and Zhao [1] was proposed a recent algorithm to distinguish spurious job post with 88 percent accuracy that is NLP and dataset-driven. Their strategy aimed at establishing linguistic patterns from job postings on semantic analysis and semantic comprehension. NLP effectiveness was revealed in the research of the study to identify suspicious posting that related to the Detection of the usage of the use of deceptive phraseology and unnatural language to identify such a post.

Johnson and Patel [2] Address feature engineering as a key determinant to identifying spam postings of jobs in terms of designing features, including posting length, recruiter activity, and metadata features. The paper shows that their model attained a 92 percent success rate and the paper showed that there is more to non-textual elements to distinguish between whether the ad is valid or fake other than considering that that it is textual in nature like user activities and structural anomalies.

Gupta and Singh [3] discussed the ethical and issues of implementing a system into operation of limiting job detection systems in authentic job places. In addition to experimentation results, they also evaluated the use of working systems that revealed a 60 percent reduction of imitations. They discussed issues of operation such as input of users, malicious activity and slowness of the system, and emphasized how the tools facilitate platform integrity and trust amongst users.

Sasidharan Pillai [4] proposed a deep learning model taking advantage of Bidirectional LSTM to verify fake job by identifying semantic patterns in job description. Information was processed forwards and backwards in the architecture, and the features of text were used in combination with other metadata including location and salary. With accuracy of 98.7 percent and AUC of 0.91, the model was successful in the identification of misleading or excess persistent language as commonly used in fraudulent job advertisements.

by Sonkar, Yadav, and R. Kumar [5] introduced a ensemble learning model to which it may incorporate Naive Bayes algorithm to perform text classification and Logistic Regression to make final prediction. The TF-IDF vectorization was applied to the system with the help of additional metadata concerning the job location or company. This methodology increased the sensitivity of detections and minimized the false demarcation, and thus it can be successfully deployed in real-time on job posting websites to better monitor fraud.

Chen, Zhao, and Lin [6] established the usage of deep learning framework for Fine detection of bogus job postings, which entails deep Convolutional network in the detection of fake job postings. They had a hierarchical Attribute derivation and achieved a very high rate in differentiation between real and fake vacancy notice using this system. The paper has shown the relevance of the current neural network architecture to train on complex textual patterns and this presents an important aspect of how diverse CNNs can be used to handle problems with high linguistic and compositional elements.

3. Proposed Methodology:

The integrative framework, on which the proposed Fake Job Requirement Prediction System is supported, contain ML and NLP, and web technologies, to detect and classify online job advertisements as real or fake. The workflow entails a multi process pipeline where a user is able to post a URL link of a job in an interface that has been constructed using MERN stack. On receiving the URL, the same is passed to the Fast API backend and the Google Safe Browsing API performs an authentication procedure to identify whether the URL is malicious or unsafe. On verification, web scraping is done to get primary job-related parameters, i.e. job title, company name, Salary, location, job type, and description. This is then taken to Gemini API where these are provided with contextualization and recognizing potential red flags as they relate to the credibility and job object of the company. All the information fed into the data processor is then introduced into an already trained ML classification model, which determines whether the listing is a fraud or not by recurring patterns in language as well as job characteristics and data source trustworthiness. Having an Admin panel that could be integrated makes the possibility of monitoring user activity all the time, activity in submission, and the classification outcome obtaining, which leads to tight control and administration. This automatic and systematic framework: facilitates the relatively quick, reliable and risk-free detection of fraudulent job ads, and therefore, it is a feasible recommendation that can protect users against internet job fraud.

3.1 Proposed Model Diagram

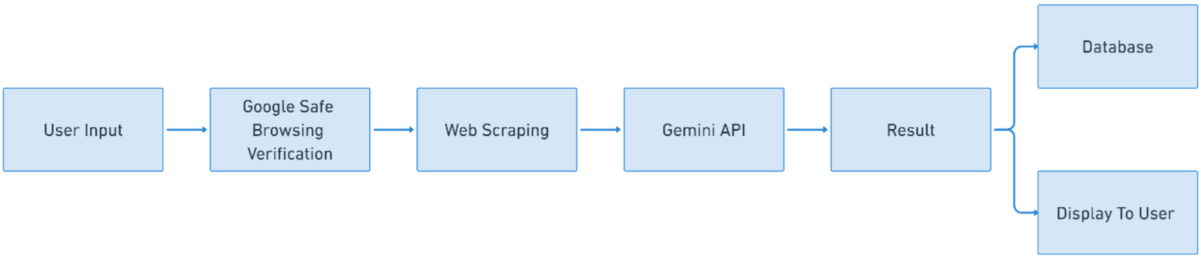


Figure 3.1.1 Proposed Diagram

Fake Job Requirement Detection System This process gets input by first submitting the job link or job details the user provides. Then, it consults Google Safe Browsing, in order to determine whether the link has been labelled as dangerous or suspicious. In case the connection is secure, the system conducts web scraping, whereby valuable information related to the job opportunity like title, description, and company details are retrieved off the webpage. The resultant extracted data is consequently analysed in the Gemini API which involves using AI methods to search indications of fake or scam job postings. According to the analysis, the system provides an output that shows whether the job is authentic or it is a fake. The outcome is stored on a database to be referred to later and the outcome is also conveyed to the user to get immediate feedback.

3.2 Block Diagram of ML Model

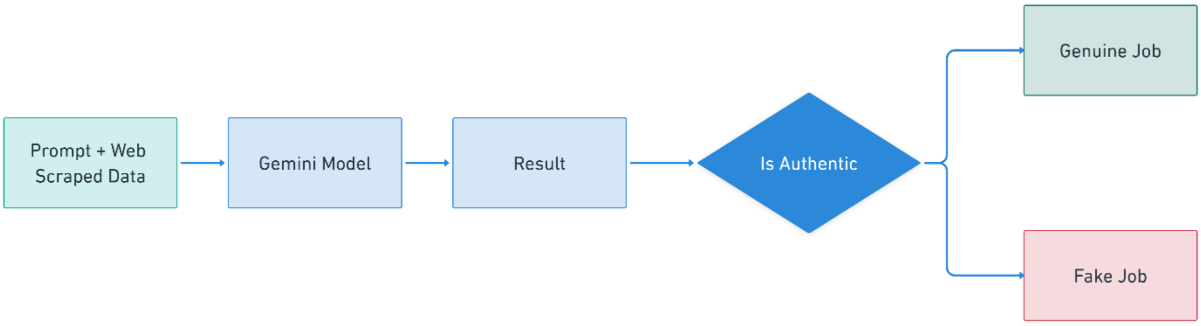
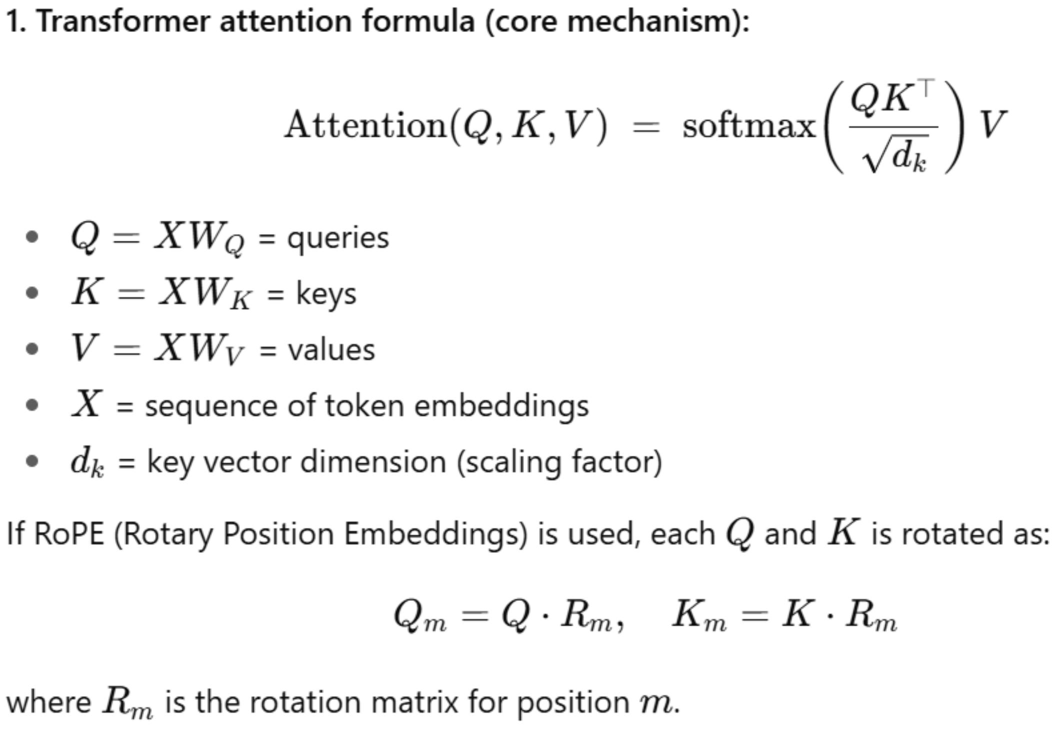


Figure 3.2.1 Block Diagram of ML Model

The mechanism of the Fake Job Requirement Detection System starts with the combination of a prepared prompt with the job information accessed by means of web scraping. This consolidated data is then feed to the Gemini AI model, which studies the information in order to detect any indications of legitimacy or fraud. The model produces an outcome and the system determines the authenticity of the job posting. According to this evaluation, some jobs are categorized as either a true job, or a phony job, and thus a user is made aware of the authenticity of the job posting.

4. Mathematical Formula:



1. Transformer attention formula (core mechanism):

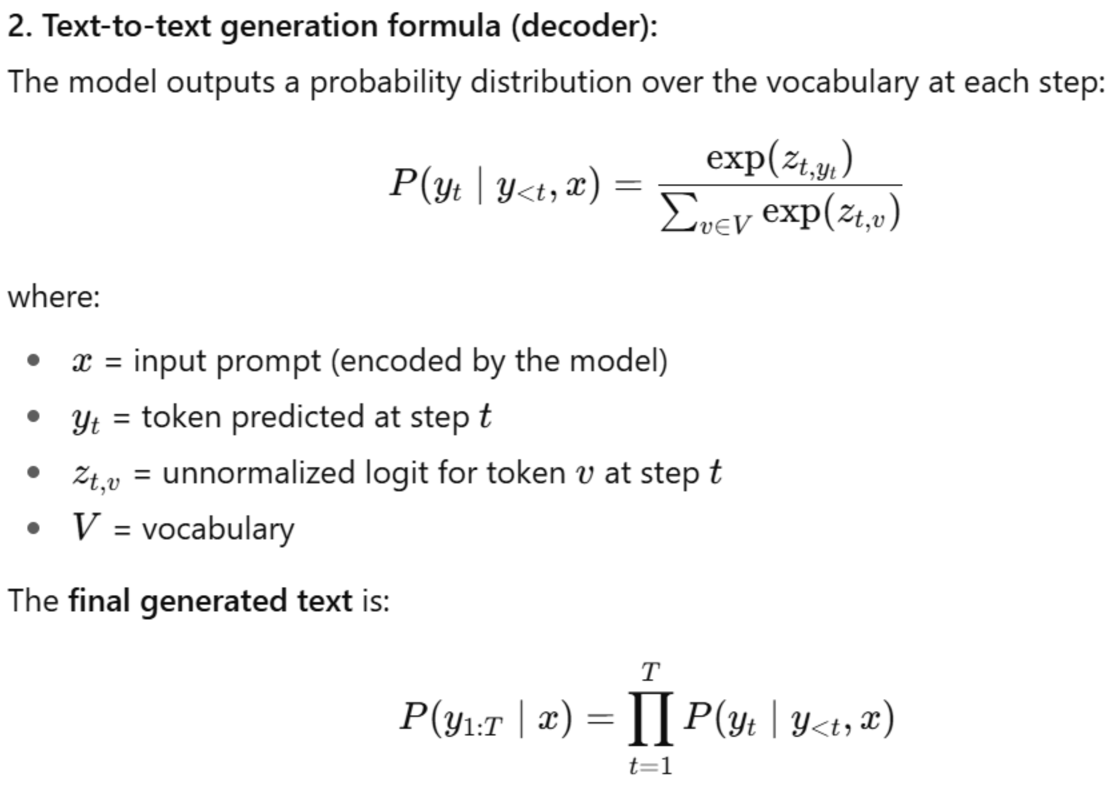
$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V$$

- $Q = XW_Q$ = queries
- $K = XW_K$ = keys
- $V = XW_V$ = values
- X = sequence of token embeddings
- d_k = key vector dimension (scaling factor)

If RoPE (Rotary Position Embeddings) is used, each Q and K is rotated as:

$$Q_m = Q \cdot R_m, \quad K_m = K \cdot R_m$$

where R_m is the rotation matrix for position m.



2. Text-to-text generation formula (decoder):

The model outputs a probability distribution over the vocabulary at each step:

$$P(y_t \mid y_{<t}, x) = \frac{\exp(z_{t,y_t})}{\sum_{v \in V} \exp(z_{t,v})}$$

where:

- x = input prompt (encoded by the model)
- y_t = token predicted at step t
- $z_{t,v}$ = unnormalized logit for token v at step t
- V = vocabulary

The **final generated text** is:

$$P(y_{1:T} \mid x) = \prod_{t=1}^{T} P(y_t \mid y_{<t}, x)$$

5. Experimental Results

S.No	Category	Test Case / Condition	Expected Behaviour	Actual Behaviour	Status
1	Job Detection	Valid LinkedIn job	Classified as real	Not Fake	Passed
2	Job Detection	Unknown portal job	Classified as suspicious	Fake	Passed
3	Job Detection	Unrealistic salary from MNC	Flagged as suspicious	Fake	Passed
4	Job Detection	Safe Browsing flagged URL	Request blocked	URL blocked	Passed
5	Job Detection	No company name	Classified using content only	Fake	Passed
6	Job Detection	Free email address used	Inconsistent source flagged	Fake	Passed
7	Job Detection	Spammy phrases (e.g., earn at home)	Detected via NLP	Fake	Passed
8	Job Detection	Official MNC careers page	Match with Gemini company data	Not Fake	Passed
9	Job Detection	Duplicate job post	Detected via semantic similarity	Fake	Passed

S.No	Category	Test Case / Condition	Expected Behaviour	Actual Behaviour	Status
10	Job Detection	Cloned job site domain	Domain flagged via Gemini	Fake	Passed
11	System Auth	Valid login	Access granted	Logged in successfully	Passed
12	System Auth	Wrong password	Show error	"Invalid Credentials" shown	Passed
13	System Auth	Register with new email	Create account	Registered successfully	Passed
14	System Auth	Duplicate email registration	Show conflict message	"Email already exists"	Passed
15	System Auth	Missing login fields	Prompt user	Form validation triggered	Passed
16	Middleware	No JWT in request	Block access	Request rejected	Passed
17	Middleware	Expired JWT token	Redirect or logout	Session expired	Passed
18	Gemini API	API temporarily down	Show fallback error	Error gracefully handled	Passed
19	Gemini API	Invalid API response	Catch and log error	No crash, fallback activated	Passed
20	Gemini API	Rate limit exceeded	Retry or message shown	Limit warning displayed	Passed
21	Web Scraping	Content not found	Skip classification	"No job data found" shown	Passed
22	Input Validation	Empty job link	Block submission	Validation message shown	Passed
23	Role Access	Admin in user route	Deny access	Access blocked	Passed
24	Network Error	During API call	Retry or error message	"Network error" shown	Passed

6. Conclusion

employment scam detection scam is now an issue of top concern throughout the glob today and the spread of scam job ads on the Internet is a significant risk to job seekers, which usually leads to a violation of privacy, monetary loss. So, we are equipped with the effect of hiring scam that is highly successful field in the research field making a huge difficulty in finding fake job posting. The present project presents a strong and smart Fake Job Requirement Prediction System, which solves this dilemma by including NLP (using Gemini API), web technology and contextual analysis. Not only does the system check the safety of job URLs posted by the user using the Google Safe Browsing API, but also automatically scrapes the necessary job-related data. The system also uses the Gemini API in order to add a contextual company information to extracted data, increasing the accuracy of classification. Constructed on the MERN stack and Fast API, the system has a convenient interface and an efficient backend that enables real-time detection and administrative control. Incorporation of new methods of natural language processing and current company intelligence and their experimental assessment show that the use of such tools increases the validity of fake job detection significantly. As a result, the solution offers a feasible and expandable way of preventing employment fraud and helps build a protected and more trustworthy recruitment environment.

7. Future Enhancement

In the light of improving the performance and using experience of the fraud Job Requirement Prediction System, a number of potential refinements are possible. The first one is the introduction of multilingual processing that would allow the platform to read job advertisements written in regional or foreign languages and thus increase its global applicability. The second improvement involves integrating such real-time adaptive response system that allow the platform to continuously improve its predictions through the use of prior misclassifications as training data. The third suggestion entails the development of a centralized reporting panel of the government officials or cybersecurity teams so that these entities can launch immediate counteraction to the known fraudulent postings. Fourth, further development of the administrator module by the use of high quality analytical tools and visualizations would assist in the constant monitoring of new trends and the detection of recruiters or areas that are of high risk. Lastly, the implementing of deep-learning methods to decode complex job postings, as well as to identifying subtitle scam patterns, may boost the capability of the platform to provide increased precision and flexibility in an ever-changing environment of remote job fraud; especially transformer-based methods or hybrid architectures.

References

[1] B. Alghamdi, F. Alharby, “An Intelligent Model for Online Recruitment Fraud Detection”, Journal of Information Security, 2019.

[2] Tin Van Huynh1, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen1, and Anh Gia-Tuan Nguyen, “Job Prediction: From Deep Neural Network Models to Applications”, RIVF International Conference on Computing and Communication Technologies (RIVF), 2020.

[3] Jiawei Zhang, Bowen Dong, Philip S. Yu, "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network", IEEE 36th International Conference on Data Engineering (ICDE), 2020.

[4] M. Sudhakar and K. P. Kaliyamurthie, "Efficient Prediction of Fake News Using Novel Ensemble Technique Based on Machine Learning Algorithm," in Information and Communication Technology for Competitive Strategies (ICTCS 2021).

[5] C. Li, G. Zhan, and Z. Li, "News Text Classification Based on Improved BiLSTM-CNN," in 2018 9th International Conference on Information Technology in Medicine and Education (ITME), 2018,

[6] Vong Anh Ho, Duong Huynh-Cong Nguyen, Danh Hoang Nguyen, Linh Thi-Van Pham, Duc-Vu Nguyen, Kiet Van Nguyen, and Ngan LuuThuy Nguyen."Emotion Recognition for Vietnamese Social Media Text", arXiv Prepr. arXiv:1911.09339, 2019.

[7] Mr. Gulshan P.1, Mr. Mukund T.2, Mr. Ajay A.3, Mr. Pankaj Kumar4, Mrs. Aruna M G5, Dr. Malatesh S H6 " Fake Job Post Prediction Using Machine Learning Algorithms".

[8] Dr. D. Madhavi1, M. Sri Manisha Reddy2, M. Ramya3, G. Sanjana4 " Detection of Online Employment Scam through Fake Jobs Using Random Forest Classifier".

[9] Dr. Y. Venkataramana Reddy , B. Sai Neeraj , K. Puneeth Reddy , P. Bhargav Reddy" ONLINE FAKE JOB ADVERT DETECTION APPLICATION USING MACHINE LEARNING".

[10] J.Raghunath , Mullah ShaikAfrin, Vemuri Bhavana, M NagaVardhan Babu, Sake Deepak, Mandadhi Avanthi , Dasari Mamatha " DETECTINGFRAUDULENTJOB POSTINGS USING MACHINELEARNING MODELS".

[11] Madhan Raj M, Mohan J, Navin Jagadish P, " FAKE JOB POSTING DETECTION".

[12] Maddi Sravya Reddy, Maddikera Hemanth Lal , Lingam Sainad, Sandeep Agarwalla " Fake Job Post Detection using Machine Learning".

[13] Ranparia D, Kumari, S., & Sahani, A. (2020). Fake Job Prediction using Sequential Network. In 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS).

[14] Chen, Y., Wang, D., & Liu, B. (2021). Fake Job Post Detection Using Convolutional Neural Networks. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval.

[15] C. Jagadeesh, Dr. Pravin R Kshirsagar, G. Sarayu, G.Gouthami and B.Manasa, "Artificial Intelligence based Fake Job Recruitment Detection Using Machine Learning Approach", Journal of Energy Sciences (JES).

[16] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu. In February 2020, Shivam BansalVersion 1 of the [Actual or Phallic] Prediction of Fake Job Posting.

[17] Wang, Z., & Liu, M. (2021). Addressing Ethical Concerns in Fake Job Posting Detection Models: A Review. IEEE International Workshop on Ethics in Artificial Intelligence.

[18] Liu, Y., & Wang, X. (2022). Fake Job Detection in Online Labor Markets: A Comparative Analysis of Algorithms. Journal of Artificial Intelligence Research.

[19] Das, S., & Chatterjee, A. (2020). Detecting Fake Job Postings using Natural Language Processing and Machine Learning. International Conference on Data Analytics.

[20] Kim, E., & Li, L. (2021). Ethical Considerations in Fake Job Posting Detection Models. IEEE Ethics in AI Symposium.