International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 1, 2026

Fraudulent Job Posting Detection

Dr.P.suresh kumar,
suresh611@gmail.com
Associate professor Department of Computer Science and Engineering, Bhoj Reddy Engineering College for Women, Hyderabad, India
Madduri Satvika,
satvikareddy080@gmail.com
Students, Department of Computer Science and Engineering, Bhoj Reddy Engineering College for Women, Hyderabad, India
Jarupula Swetha,
jarupulaswetha19@gmail.com
Students, Department of Computer Science and Engineering, Bhoj Reddy Engineering College for Women, Hyderabad, India
Jaithwaram Varsha,
1202varsha@gmail.com
Students, Department of Computer Science and Engineering, Bhoj Reddy Engineering College for Women, Hyderabad, India

Abstract:
The rapid growth of online job portals has increased accessibility for job seekers, but it has also led to a rise in fraudulent job postings. These fake listings often aim to deceive users by collecting personal information or money. This project presents a machine learning-based approach to detect fraudulent job postings using Natural Language Processing (NLP) techniques. The system uses TF-IDF (Term Frequency–Inverse Document Frequency) for feature extraction and classification algorithms such as Naive Bayes and Support Vector Machine (SVM) to classify job postings as real or fake. The proposed system improves detection accuracy, reduces manual effort, and enhances user safety in online recruitment platforms.

Keywords: *Machine Learning, TF-IDF, Naive Bayes, SVM, Fraud Detection, NLP*

1 Introduction

The increasing use of online recruitment platforms has transformed the job search process, making it faster and more convenient. However, this growth has also led to a rise in fraudulent job postings that mislead applicants by offering fake opportunities. These scams often result in financial loss and misuse of personal data.

Traditional detection methods rely on manual verification or rule-based systems, which are inefficient and not scalable. As fraudulent techniques become more sophisticated, there is a need for an automated and intelligent system to detect such postings.

This project proposes a machine learning-based system that analyzes job descriptions and classifies them as real or fake using NLP techniques. By automating the detection process, the system aims to improve reliability and protect users from online job scams.

2 Literature Survey

Recent research has focused on using machine learning and NLP techniques to detect fraudulent job postings and online scams. Various supervised learning algorithms have been applied to classify job listings based on textual features.

Some studies have used deep learning models to improve detection accuracy by capturing complex patterns in data. However, these approaches often require high computational resources and large datasets.

Other research has shown that traditional machine learning algorithms like Naive Bayes and Support Vector Machine (SVM), combined with TF-IDF feature extraction, provide efficient and accurate results for text classification tasks.

Despite these advancements, challenges such as scalability, data imbalance, and evolving fraud patterns still exist. The proposed system addresses these issues by using a combination of efficient algorithms and feature extraction techniques to achieve better performance.

3 Methodology

The proposed system adopts a machine learning-based approach to detect fraudulent job postings by integrating data preprocessing, feature extraction, and classification techniques. The system architecture is designed to process textual job data and classify it into genuine or fraudulent categories with high accuracy.

The dataset used in this study is obtained from Kaggle and consists of labeled job postings containing

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

features such as job descriptions, titles, and company-related information. This dataset provides a real-world representation of recruitment data, enabling effective training and evaluation of machine learning models.

Initially, the collected data undergoes preprocessing to remove noise, stop words, and irrelevant characters. This step ensures that the textual content is cleaned and standardized for further analysis. Following preprocessing, Term Frequency–Inverse Document Frequency (TF-IDF) is applied to convert textual data into numerical vectors, allowing machine learning algorithms to interpret the data effectively.

The system incorporates classification algorithms such as Naive Bayes and Support Vector Machine (SVM) to analyze the extracted features. Naive Bayes predicts the probability of a job posting being fraudulent based on word distributions, while SVM identifies an optimal boundary that separates real and fake job postings.

The integration of dataset-driven learning with feature extraction techniques enables the system to identify hidden patterns in job descriptions and improve prediction accuracy. This methodological approach ensures scalability, efficiency, and reliability in detecting fraudulent job postings

4 Implementation and Experimental Setup

The implementation of the proposed system is carried out using a Python-based environment supported by standard machine learning libraries. The system operates in a web-based interface where users can input job descriptions and obtain prediction results.

The dataset collected from Kaggle is divided into training and testing subsets to evaluate model performance under realistic conditions. During the training phase, machine learning models such as Naive Bayes and Support Vector Machine (SVM) learn patterns from the dataset and adjust their parameters accordingly. In the testing phase, the trained models are evaluated using unseen data to measure accuracy and consistency.

The system processes input data by applying TF-IDF feature extraction and passing the transformed data to the trained models for prediction. The output is then displayed to the user, indicating whether the job posting is real or fraudulent.

The experimental setup includes multiple test scenarios involving different job descriptions to analyze system performance, response time, and prediction accuracy. The results demonstrate that the system effectively identifies fraudulent job postings while maintaining usability and efficiency.

Furthermore, the use of a real-world dataset enhances the reliability and practical applicability of the system. The integration of machine learning techniques with efficient data processing ensures that the system can handle large volumes of job postings and provide accurate predictions in real-time environments.

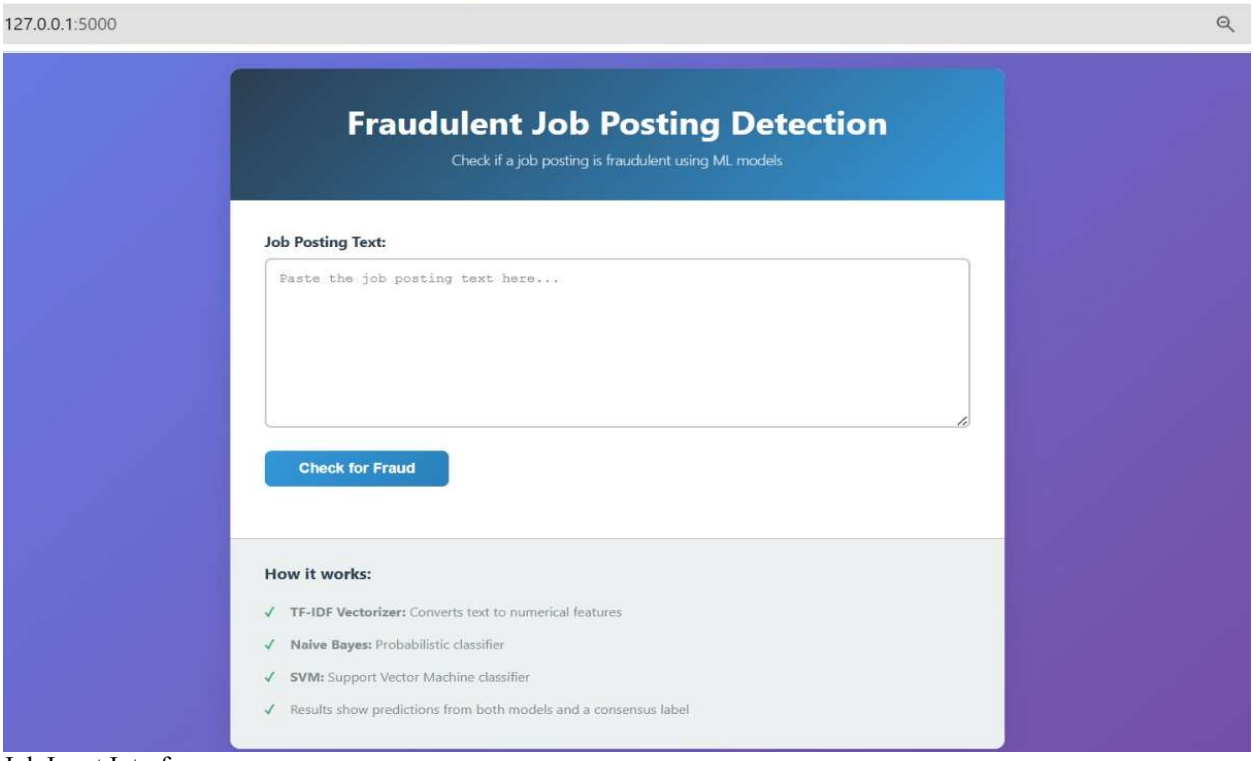
5 dataset

Fake job postings dataset

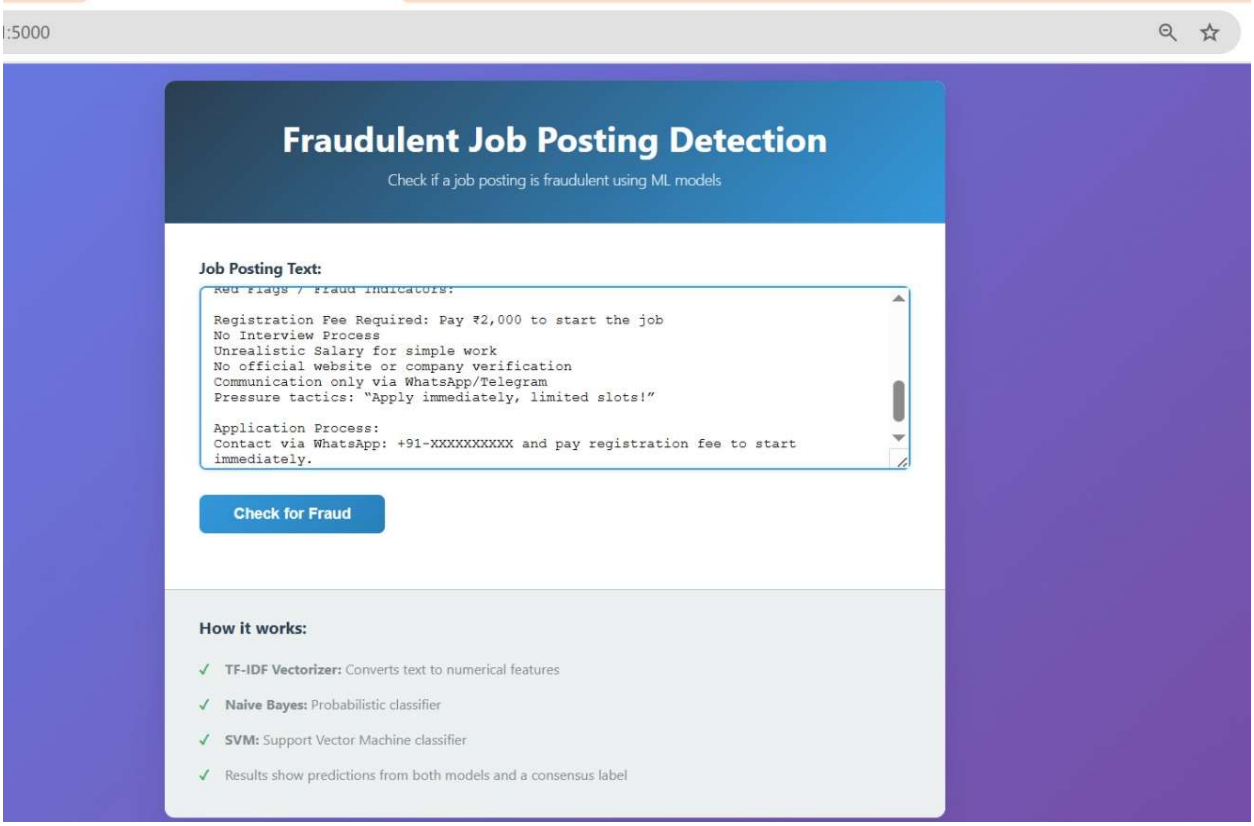
id	title	location	department	salary_rang	company_p	description	requiremen	benefits	telecommu	has_compa	has_questio	employmen	required_ex	required_ed	industry	function	fraudu
1	Marketing I	US, NY, New	Marketing		We're Food	Food52, a fa	Experience with content		0	1	0	Other	Internship			Marketing	
2	Customer Si	NZ, , Auckla	Success		90 Seconds,	Organised -	What we ex	What you w	0	1	0	Full-time	Not Applicable		Marketing a	Customer Si	
3	Commissior	US, IA, Wever			Valor Servic	Our client, l	Implement pre-commiss		0	1	0						
4	Account Exe	US, DC, Was	Sales		Our passion	THE COMPA	EDUCATION	Our culture	0	1	0	Full-time	Mid-Senior l	Bachelor's D	Computer S	Sales	
5	Bill Review I	US, FL, Fort Worth			SpotSource	JOB TITLE: I	QUALIFICAT	Full Benefits	0	1	1	Full-time	Mid-Senior l	Bachelor's D	Hospital & H	Health Care	
6	Accounting	US, MD,				Job Overview	Apex is an environment		0	0	0						
7	Head of Cor	DE, BE, Berli	ANDROIDPI	20000-2800	Founded in	Your Respo	Your Know-	Your Benefi	0	1	1	Full-time	Mid-Senior l	Master's De	Online Medi	Manageme	
8	Lead Guest	US, CA, San Francisco			Airenvy's mi	Who is Aire	Experience	Competitive	0	1	1						
9	HP BSM SM	US, FL, Pensacola			Solutions3 i	Implementa	MUST BE A US CITIZEN./		0	1	1	Full-time	Associate		Information	Technology	
10	Customer Si	US, AZ, Phoenix			Novitex Ent	The Custom	Minimum Requirements		0	1	0	Part-time	Entry level	High School	Financial Se	Customer Si	
11	ASP.net Dev	US, NJ, Jersey City		100000-120000		Position : #U	Position : #U	Benefits - Fu	0	0	0	Full-time	Mid-Senior l	Bachelor's D	Information	Information	
12	Talent Sourc	GB, LND, Lo	HR			Want to bui	TransferWis	We're looki	You will join	0	1	0					
13	Applications	US, CT, Stamford			Novitex Ent	The Applica	Requirements:4 – 5 year		0	1	0	Full-time	Associate	Bachelor's D	Manageme	Information	
14	Installers	US, FL, Orlando			Growing eve	Event Indus	Valid driver's license,Sor		0	1	1	Full-time	Not Applica	Unspecified	Events Servi	Other	
15	Account Exe	AU, NSW, Sy	Sales		Adthena is t	Are you inte	You'll need	In return we	0	1	0	Full-time	Associate	Bachelor's D	Internet	Sales	
16	VP of Sales	SG, 01, Sing	Sales	120000-150	Jungle Ventı	About Vault	Key Superpc	Basic: SGD 1	0	1	1	Full-time	Executive	Bachelor's D	Facilities Se	Sales	
17	Hands-On C	IL, , Tel Aviv	R&D		At HoneyBo	We are look	Previous experience in c		0	1	0	Full-time	Mid-Senior l	level	Internet	Engineering	
18	Southend-o	GB, SOS, Southend-on-Sea			Established	Governmen	16-18 year (	Career pros	0	1	1						
19	Visual Desig	US, NY, New York			Kettle is an i	Kettle is hiring a Visual Designer!Job			0	1	0						
20	Process Cor	US, PA, USA Northeast			We Provide	Experienced	Must have 5 or more yea		0	0	0	Full-time					
21	Marketing A	US, TX, Austin			IntelliBright	IntelliBright	Job RequirementsAssist		0	1	0					Marketing	
22	Front End D	NZ, N, Auckland			Frustrated v	Want to be	You will mo	You will be p	0	1	0	Full-time	Mid-Senior l	Master's De	Consumer E	Engineering	
23	Engagemen	AE, ,	Engagement		Upstream's	The positior	Requiremer	Salary &amp	0	1	1	Full-time	Mid-Senior l	Bachelor's D	Telecommu	Sales	
24	Vice Preside	US, CA, Carl	Businessfrie	100000-120	WDM Grou	#URL_eda2	Job Require	Businessfrie	0	1	0	Full-time	Executive	Unspecified	Internet	Sales	
25	Customer Si	GB, LND, London				We are a canary wharf based e-com			0	0	0						
26	H1B SPONS	US, NY, New York			i28 Technol	Hello,Wish y	JAVA, .NET, SQL, ORACLI		0	1	1						
27	Marketing E	SG, ,	Marketing		If working ir	We are curr	This positio	We are look	0	1	0	Full-time	Associate		Online Medi	Marketing	
28	HAAD/DHA	AE, AZ, Abu	Medical		We the Mec	HAAD/DHA	Requiremer	Our client is	0	1	0	Full-time	Associate	Master's De	Hospital & H	Health Care	

6 output

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

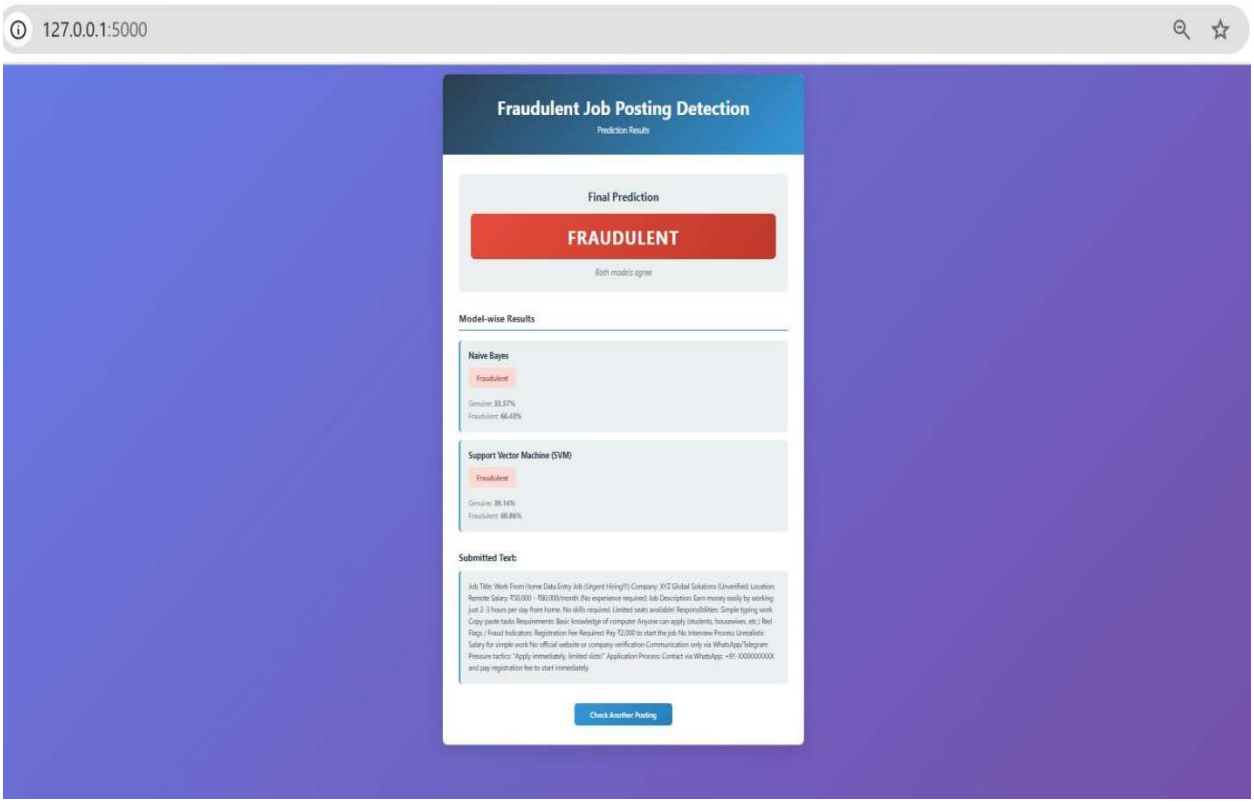


Job Input Interface

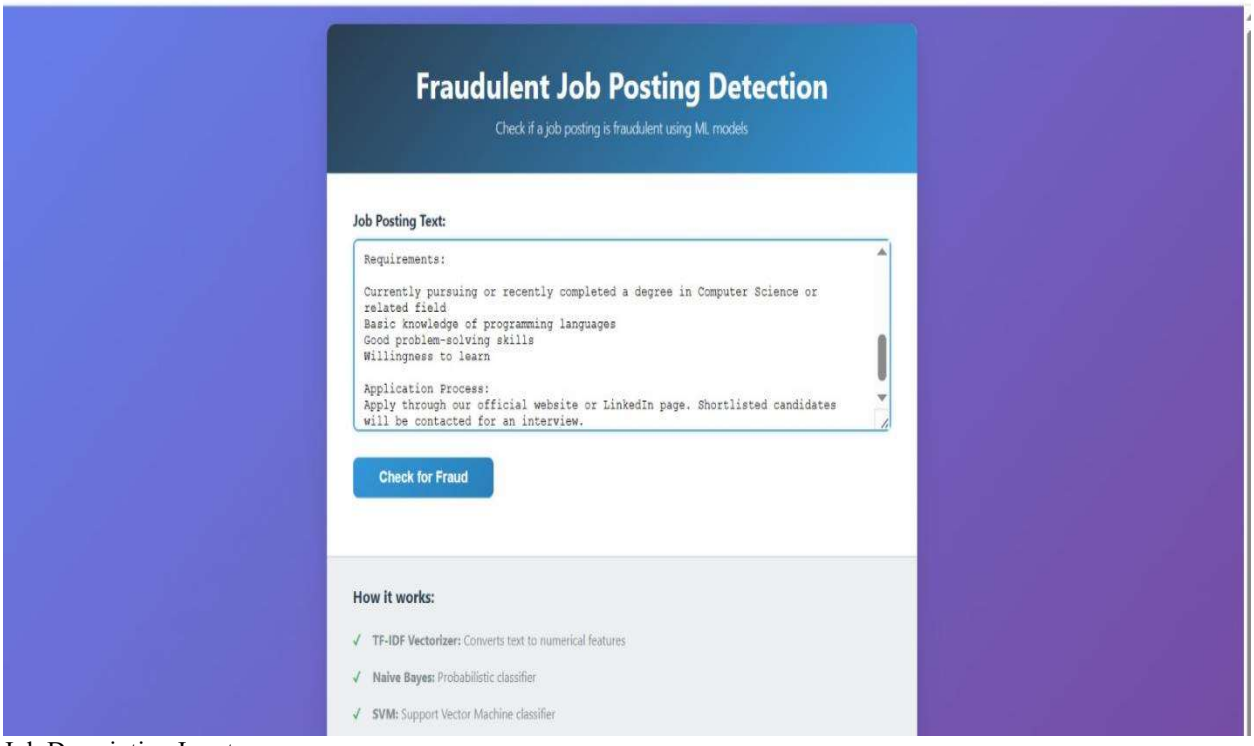


Job Description Input

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

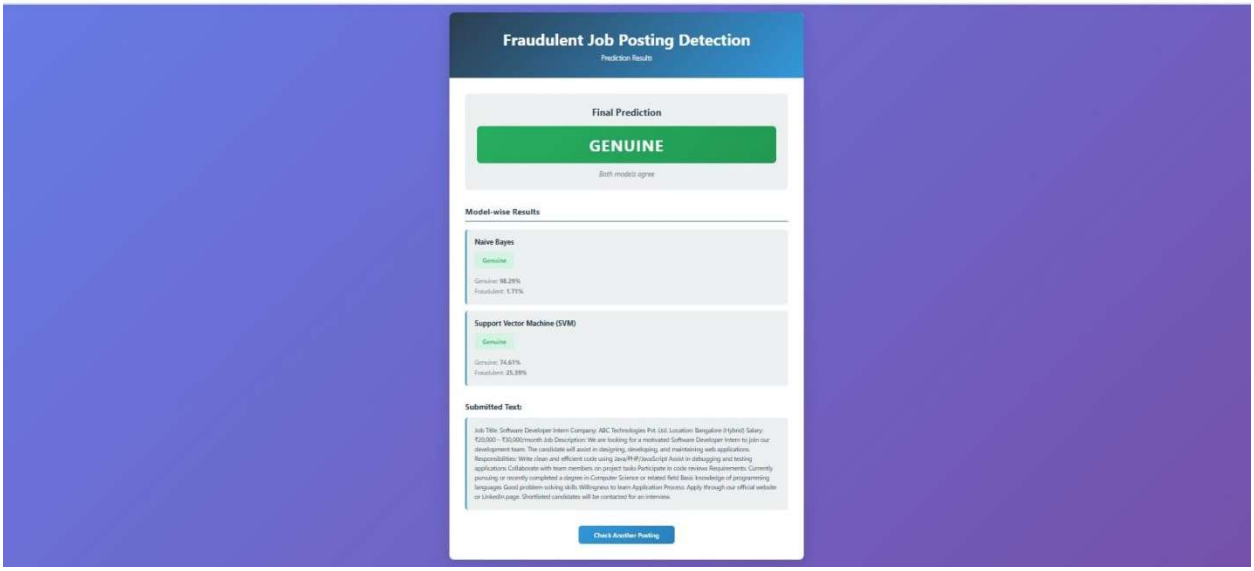


Prediction Result



Job Description Input

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026



Prediction Result

7

Conclusion

This paper presented a machine learning-based approach for detecting fraudulent job postings using Natural Language Processing techniques. The proposed system utilizes TF-IDF for feature extraction and classification algorithms such as Naive Bayes and Support Vector Machine to analyze job descriptions and identify fraudulent patterns.

The implementation demonstrates that the system is capable of accurately classifying job postings as genuine or fraudulent based on textual features. By automating the detection process, the system reduces manual effort and enhances the safety of job seekers on online recruitment platforms. The use of a real-world dataset further improves the reliability and practical applicability of the system.

Although the system achieves effective performance, certain limitations such as dependency on dataset quality and evolving fraud patterns may affect accuracy. Future work can focus on integrating advanced deep learning techniques, real-time data processing, and improving model robustness to handle complex fraud scenarios.

Overall, the proposed system provides a scalable and efficient solution for detecting fraudulent job postings and contributes toward building a safer and more trustworthy online job environment.

8 **References**

1. D. H. Vu, “Improving Fake Job Detection Using Deep Learning-Based NLP Techniques,” *Journal of Applied Artificial Intelligence*, 2025.
2. T. N. Goud, “A Machine Learning Approach for Detecting Fraudulent Job Postings,” *Journal of Smart Internet and Information Systems*, 2025.
3. A. S. Filani, “Development of a Fake Job Posting Detection System Using Machine Learning,” 2025.
4. “Fake Job Detection and Analysis Using Machine Learning Techniques,” *IJRPR Journal*, 2025.
5. “Fake Job Detection Using Logistic Regression and TF-IDF,” *IJFMR Journal*, 2025.
6. “Fake/Real Job Posting Detection Using Machine Learning,” *IJRASET*, 2024.
7. “Fake Job Recruitment Detection Using Machine Learning Techniques,” *IJNRD*, 2024.
8. S. Kondapalli, M. Dudala, K. K. Kumar, K. Spurthi, K. S. Kumar and R. D. Kumar, "Deep Learning Convolutional Nets: Intelligent System for Paddy Leaf Disease Diagnosis," 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON), Tumkur, India, 2025, pp. 1-8, doi: 10.1109/SSITCON66133.2025.11341943
9. Y. Sowjanya, S. Gopalakrishnan and R. D. Kumar, "Internet of Things in Health Care: Motivation and Challenges: A Survey," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-7, doi: 10.1109/ICCCNT61001.2024.10725769.
10. Vojja, L., Kumar, R.D., Sivaprasad, P.V.S., Satyanarayana, A. (2025). Encryption with Identity Based Approach for Versatile Encrypted Data Sharing in Public Cloud. In: Farhaoui, Y., Herawan, T., Lucky Imoize, A., Allaoui, A.E. (eds) Intersection of

Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment. ICAISE 2024. Lecture Notes in Networks and Systems, vol 1397. Springer, Cham. https://doi.org/10.1007/978-3-031-90921-4_81