SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

DATA INTELLIGENCE FOR IDENTIFYING FRAUDULENT JOB OPPORTUNITIES

Amit Garg[1], Gourav Pal[2], Ameer Alam[3]
[1]*Assistant Professor, KCCITM, Greater Noida, India*
[2]*Student, KCCITM, Greater Noida, India*
[3]*Student, KCCITM, Greater Noida, India*

Article DOI: https://doi.org/10.36713/epra25634
DOI No: 10.36713/epra25634

ABSTRACT

In today's digital era, most people look for job opportunities through online job portals and recruitment websites. However, fake job postings have become a major concern for both job seekers and employers. These fake posts waste valuable time and can even lead to scams or data theft.

To address this issue, our project aims to build a machine learning model that can automatically detect whether a job posting is real or fake. Using data mining and classification techniques, the proposed system helps in identifying fraudulent job advertisements efficiently and accurately.

I. INTRODUCTION

Data Intelligence For Identifying Fraudulent Job Opportunities is web page for verify that the job post is fraud or real to save time of applicantce and keep them prevented from fack jobs and scams like demanding money as a registration and ask them to circulate their job poster or post in their Institutes group. Fake job ads don't just drain an applicant's time and energy they can open the door to scams and steal personal details, like a home address or phone number.

This study tackles the problem by creating a machine learning system that can detect fake job postings automatically like indicate an ad that promises "free luxury travel" with no real company behind it. The model uses data mining and classification to detect fraudulent listings quickly, like catching a fake job post before it publishes, helping keep online recruitment safer and more trustworthy for employers and job seekers.

Modern detection techniques is Integration of Natural Lanaguage Processing(NPL),Machine Learning Algorithms, and Data Mining for analyzing different techniques of job Description,as similar salary offer by faraud job,there can too much grammer error as we see in fack message scams or doubtful contact information.So,We can train our system on these kind of labled datasets which contains fake job posts and we should temp mobile number or email not use personal email or mobile number.Online job posting and hiring improving fraudulent job posting to development of protection system of fraudulent job.

Techniques like random forest and Support Vector Machine(SVM).Points like job description words and length ,wrong company details,SVM and Random Forest can get approx 95% of accuracy in detection of fraulent job posting.NPL is helpful to understand semantic structure of job posts to Identifying Fraudulent Job Opportunities.

Aspects like salary,location of the job is constant and also a pattern of job can help to understand the job seems real or fake.
Recent techniques also acquire techniques of real –time detection.Scammers are always be updated to do fraud, so we need fresh data to reamin effective.Leading job marketplaces like LinkedIn,Naukri.com,Indeed.com and Shine.com these machine learning systems heps to reduce faud posting,increasing trust and precaution for users.Identifying Fraudulent Job Opportunitities not only for prevent from fraud jobs but also make more healthier to digital job market in which job seekers will only see real jobs.

II. PROBLEM STATEMENT

The beginning of online job postings, discriminating real job openings from fraud employment is increasingly difficult. The fake volume of online data makes the manual detection of fraud openings Unfortunately inefficient. This calls for an automate, data driven approach for identifying fake online job postings. An automated classification provide fraud protection to job seekers, boosts user confidence, and increase the trustworthiness of online employment sites.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

In recent year Fraudulent job portals have become most common and easy way to Cheat.The fraudulent record are often generated with bad intentions like stealing someone's personal details to cheat for money,and send fake messages to cheat them or get private information with the person permission.As misleading posts become sophisticated.

One of the major challenges in detection of these posts is details which scammers gives seems genuine and they advertise often then seekers watch that and they feels like the structure ,formate and language creates difficult to differentiate in first time.More than they publish it in large volume of job with is people harms themselves with situational harm ,financial loss or data misuse.
Data mining and machine learning methods are useful for identify fake job advertisement and accuracy depends on datasets,many existing studies points on only single algorithm at a time but there is lack of research with comparision of different data mining techniques.Important things verify real job includes clearity of job title,salary consistency,company verification,style of witing and pattern of posting or structure of detail.

The goal of research to make safe online job platform by providing a clear comparision of different data mining strategy and we choose most effective tricks to reduce fake job postings and protect job seekers from scams.

Related Work:-
There is a number of works on the detection of fake jobs as well as in related topics such as detection of fake news and spam mails. Most of these papers use machine learning algorithms for detection offake jobs from a pool of real jobs. Some of the common methodology used in various research papers for the detection of fake jobs are:
- Using publicly available dataset for the detection of fake jobs.
- Using machine learning algorithms in order to classify the fake jobs and real jobs from a pool of job posts
- Using ensemble classification modeling to increase the accuracy of machine learning algorithms[1].

For classification, supervised and unsupervised, both algorithms can work. Random forest agave learning-based approach where each classifier comprehends numerous tree-like classifiers applied to various examples, and each tree votes in favor of the most fitting class. Another helpful technique can be boosting, which can work with multiple classifiers for a single classifier to improve classification results. Extended innovation applies an algorithm for classifying the weighted adaptations of training data and chooses the grouping of the more significant voting classifier. AdaBoost illustrates a procedure of boosting, which delivers better effectiveness (Murtagh, 1991). Expanding algorithms implies tackling issues with spam filtration viably. In addition, Gradient boosting is an extra boosting procedure for a Classifier dependent on the decision tree rule (Prentzas et al., 2019). It likewise limits the deficiency of accuracy[2].
A common method for text-analysis and text-based classification is to process for term-frequency or patterns of terms. However, these features alone may not be able to differentiate fake and authentic job advertisements. Thus, in this work, we proposed a method to detect fake job recruitments using a novel set of features designed to reflect the behavior of fraudsters who present fake information. The features were missing information, exaggeration, and credibility. The features were designed to represent in the form of a category and an automatically generatable score of readability. Data from EMSCAD dataset were transformed in accordance with the designed features and used to train a detection model for fake job detection. The experimental results showed that the model from the designed features performed better than those based on the term-frequency approach in every applied machine learning technique. The proposed method yielded 97.64% accuracy, 0.97 precision and 0.99 recall score for its best model when used for classifying fake job advertisements[9].The number of job openings advertised on the internet has also increased. It takes time and effort to apply for a job online. It may also cause concern because they have access to our personal, professional, and academic data. As a result, it is critical to determine whether the job we are looking for is legitimate or not. This model is being developed to detect bogus job postings. We employed Deep Neural Network in this project. For comparison, Machine learning algorithms like K Nearest Neighbour, NB classifier, decision tree, SVM, and RF classifiers were also used in this work.[12]

The Employment Scam Aegean Dataset was utilized to build Machine Learning classification models using Logistic Regression, Support Vector Machine, Decision Tree, and Naïve Bayes algorithms.These models were combined with different vectorizers like Term Frequency-Inverse Document Frequency, Bag of Words, and Hash.This model shows promise in effectively identifying fraudulent job advertisements, safeguarding job seekers from scams in the online job market.[10]

Spammers can manipulate reviews for gaining profit and hence it is required to develop techniques that detects these spam reviews. This can be implemented by extracting features from the reviews by extracting features using Natural Language Processing (NLP). Next, machine learning techniques are applied on these features. Lexicon based approaches may be one alternative to machine learning techniques that uses dictionary or corpus to eliminate spam reviews[3].

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

III. OBJECTIVES

The exteme growth of online recruitment platforms has created new opportunities for employe and seeker.It opened opportunities to grow number of fraud jobs that cheats,steal data.these fraudulent things continuous increase number of fake jobs,this study helps to examining different data mining techniques to detect fraudulent jobs.

Our next objective is to improve classification of results based on differentiation of data categories with exploring different combination of text , number and metadata features,the study looks for quality and selection property like job title clarity,company credibility,frequency of posting,grammatical consistency and salary pattern .the objective is to not to detect current features but also to new enhaces pattern in future.

The next objective is to compare understandability and practical usefulness of data mining aspects.Some algorithms offer high accuracy,but they may be difficult to explain or implement in real-world systems.This part of study evaluates which models are more suitable for real –world deployment on job sites.

Finally ,our aim is to build a effective fraudulent job identification system based on detecting.These recommendations can guide developers ,researchers and cybersecurity teams ti create safe online job portals.
a)To study and compare the performance of different data mining and machine learning techniques for detecting fake job posts.
b) To implement and analyze the Random Forest Classifier and K-Nearest Neighbors algorithms for prediction.
c)To design and develop a website for real-time fraud job identification.
d)To make a reliable and high accurate system that helps job seekers identify real opportunity and prevent from scams.

IV. LITERATURE REVIEW

The growth of online job portals has come with growth of fake job posts, creating a challenge for users to try valid job offers from fraudulent. Manually identifying fake job posts is a slow and unreliable process and our current system cannot effectively manage the ever increasing volume of online traffic. The development of an automation system that could accurately assess job postings and identify fraudulent ones would be beneficial. It would minimize the risk of users being scammed, increase the system trustwortiness, and reduce the amount of work for human research.

As more people coming going online for job,researchers have start develop platforms to build for detect fake posts or ads automatically.Most of the student just try to identify how a scam ad looks.people get that the these fauds conatain unusual words,to remain popular about job,promise salary that’s unexpected,or don’t give company information.these sign’s help researchers to choose which aspect use train Machine.

With the passing of years,Machine learning techniques are becoming popular.To get clear cut on the bases of text uses like Logistic Regression,Naïve Bayes and Support Vector Machines use too much .As field moved decision- tree models getting more used .random forest because it can handle noisy data and avoids overfitting better than useual model.

A beggest problem for seekers is that the how we can compare scam advertisement with real.To prevent with this we can try oversampling,undersampling or both and also we can look into language details –how many times a word showen,the tone of advertisement written or spoken,how it silar to real on or readale like that ,complication of grammer.other aspects of structure salary ranges,company info,job category,dates.there’s n’t any perfect rule which is fixed to identify in future fraudulent job posts.

Futuristic Enhancement:-
Expand the system to operate across multiple major job platforms such as Indeed, Glassdoor, and Monster while simultaneously incorporating user behavior analytics. By analyzing user interactions,including clicks,application patterns,and engagement trend,the system can identify suspicious posting characteristics and detect tendencies that indicate fake or high-risk job opportunities more accurately[4].

V. METHODOLOGY

Datasets:-This research works on a dataset from Kaggle to categorize a job advertisement as fraudulent or not based on some attributes derived from the advertisements available on different sources. The data was available in a CSV file having 1000 instances of jobs advertisements. Each advertisement is defined in terms of attributes on which we are working, that data is then preprocessed and classified through several algorithms. As the dataset has many missing values and anomalies, it needs a preprocessing step before it can be used as an input to any classification algorithm.and for processing of datasets the initial dataset had -- attributes based on which this

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

model would be predicting the status of an advertisement.At present, many machine learning algorithms have been used to detect such fraudulent posts. But, the performance of such algorithms to be measured and compared to find a proper algorithm to incorporate in identifying fake things. In this research, the use of a proposed model with the help of Microsoft Azure Machine Learning Studio tested a comparison study on the performance of a two-class boosted decision tree and two-class decision forest algorithms. Researchers used F1 Score, Recall. Accuracy and precision to compare those two algorithms. Results showed that a two-class boosted decision tree is better for detecting fake job posts than the two-class forest decision algorithm. Thus, a two-class decision forest algorithm can be used to find and identify false or gossip messages, tweets, and social media publications[8]. These -- attributes include job id, title, location, department, salary range, company profile, description, requirements, benefits, and telecommuting, has the company logo, has questions, employment type, required experience, required education, industry, and function. Each attribute contains either object or integer data[2].

Fig. 1 shows the distribution of fake and legitimate job advertisements in the dataset.

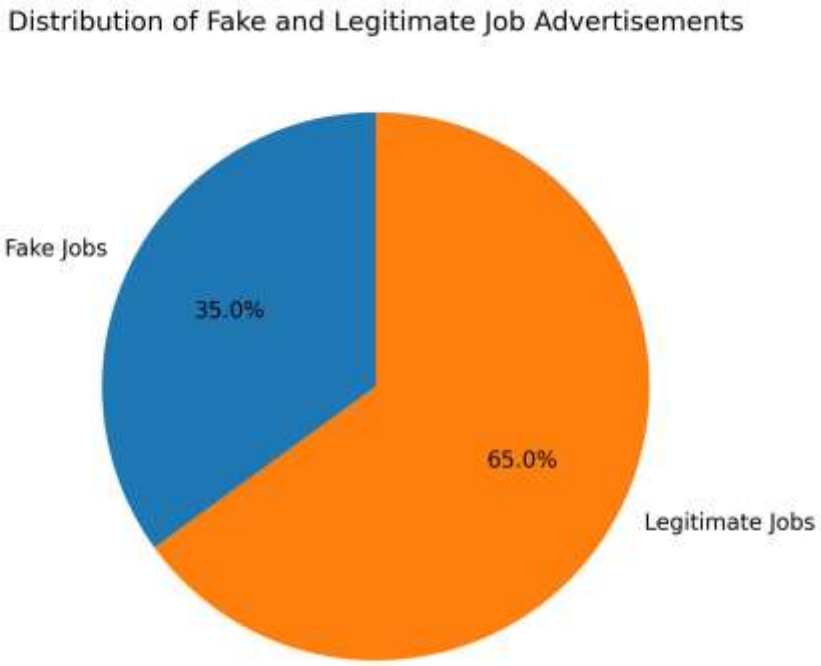


Fig. 1. Distribution of fake and legitimate job advertisements

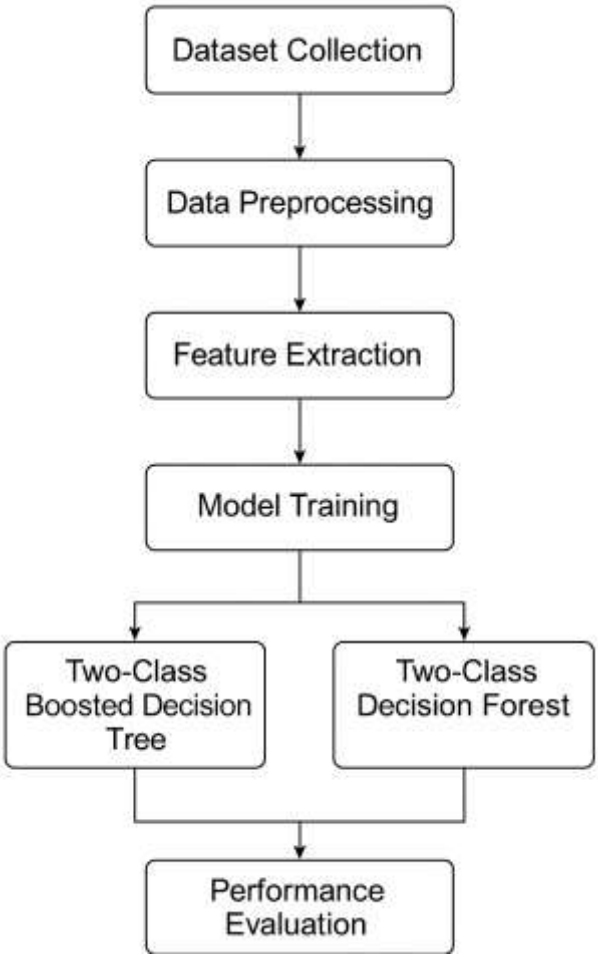
While evaluating performance skill of a model, it is necessary to employ some metrics to justify the evaluation. For this purpose, following metrics are taken into consideration in order to identify the best relevant problem-solving approach. Accuracy is a metric that identifies the ratio of true predictions over the total number of instances considered. However, the accuracy may not be enough metric for evaluating model‘s performance since it does not consider wrong predicted cases. If a fake post is treated as a true one, it creates a significant problem. Hence, it is necessary to consider false positive and false negative cases that compensate to misclassification. For measuring this compensation, precision and recall is quite necessary to be considered[3]. 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.[5]

Like many other classification tasks, fake job posing prediction leaves a lot of challenges to face. This paper proposed to use different data mining techniques and classification algorithm like KNN, decision tree, support vector machine, naive bayes classifier, random forest classifier, multilayer perceptron and deep neural network to predict a job post if it is real or fraudulent. We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples. Deep neural network as a classifier, performs great for this

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

classification task. We have used three dense layers for this deep neural network classifier. The trained classifier shows approximately 98% classification accuracy (DNN) to predict a fraudulent job post.[7]

Random Forest is an ensemble learning method that constructs multiple decision trees during training and aggregates their outputs for classification tasks. It is known for its ability to handle large datasets with higher dimensionality. In this project, the Random Forest model was configured with multiple decision trees. This setup allows the model to make robust predictions based on the collective output of multiple trees, thereby enhancing its accuracy in classifying job posts[6].By analyzing job descriptions and recruiter details using deep learning techniques specifically, natural language processing (NLP) and neural networks this study seeks to identify such scams. The solution improves job site security by spotting suspicious trends, which helps job seekers avoid fraud and makes the online hiring process safer.[10]



A. Dataset Collection

Dataset for this study came from Kaggle, while including job ads marked either fake or real. It's saved as a CSV file with around 1000 entries. Every listing has several features pulled from various job websites instead of just one source.

B. Dataset Attributes

The data's made up of categories (like text) but also numbers (whole digits). What matters most here is:

- Job ID
- Job Title
- Location
- Department
- Salary Range

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

- Company Profile
- Job Description
- Requirements
- Benefits
- Telecommuting
- Company Logo Availability
- Screening Questions
- Employment Type
- Required Experience
- Required Education
- Industry
- Job Function

These traits act as clues to guess if a
job ad's fake or real - using them helps spot scams before you click. Each detail gives a hint, so together they build the full picture wit hout needing jargon or hype.

C. Data Preprocessing

Thedataset has gaps, errors, or odd entries - so it needs cleaning before sorting info. Cleaning means fixing these issues first - then the data works better for grouping tasks

1. Dealing with gaps or empty spots
2. Cleaning out cluttered or pointless info
3. Encoding categorical variables
4. Normalizing numerical attributes
5. Getting the tidy data ready for ML tools

This move helps the model work smarter while staying steady - also making results more trustworthy through consistent tweaks now a nd then.

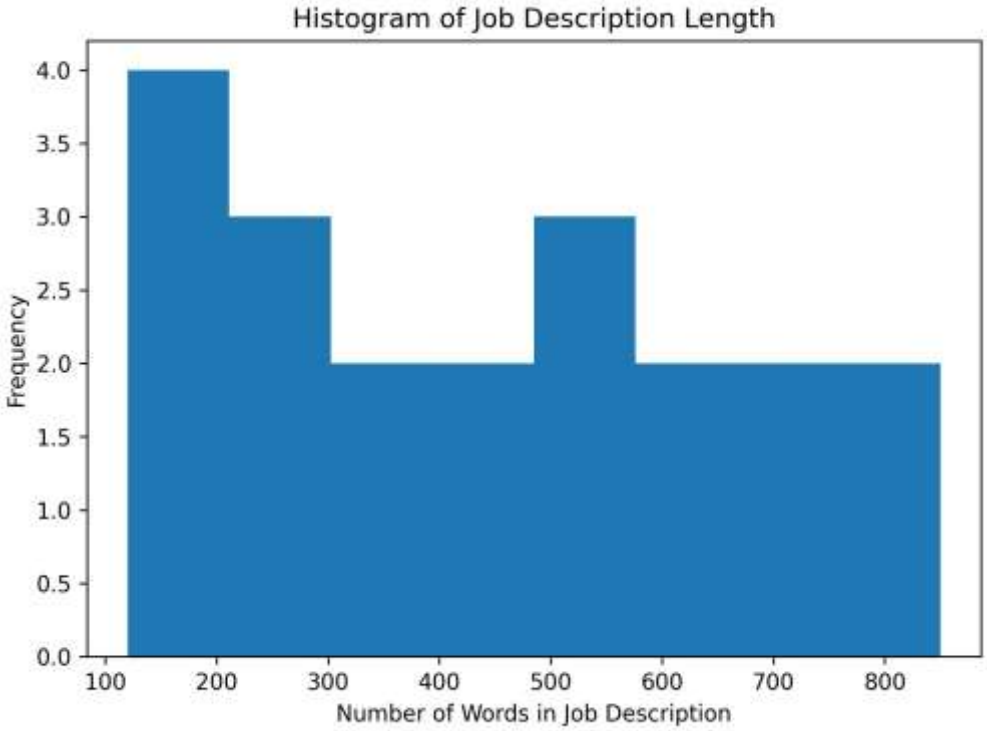


Fig. 2 illustrates the distribution of job description lengths used for text-based feature extraction.

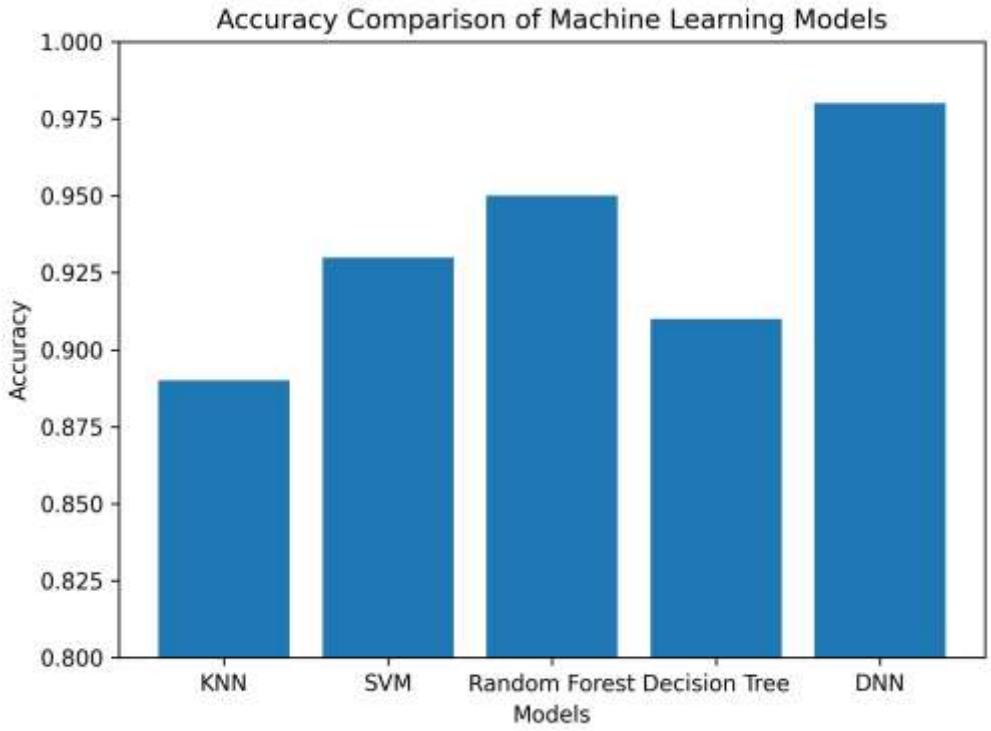
SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

D. Picking models plus sorting methods

Earlier work looked into different **datamining plus machine learning** methods to spot fake job ads. This study compared them using **Microsoft Azure Machine Learning Studio** by testing several approaches side by side

- Two-Class Boosted Decision Tree
- Two-Class Decision Forest

  Also, some methods mentioned in earlier studies are:
- K-Nearest Neighbors (KNN)
- Decision Tree
- Support Vector Machine (SVM)
- Naïve Bayes Classifier
- Random Forest
- Multilayer Perceptron (MLP)
- **Deep Neural Networks (DNN)**



The accuracy comparison of different machine learning models is presented in Fig. 3.

E. How Well It Works Measures

To check how well the models worked, we looked at several different measures:

- **Accuracy** checks how many predictions were right compared to all tries
- **Precision** tells you the share of flagged job listings that truly are scams
- Recalls show the number of real scam listings that actually get caught
- **F1-Score**: It's a balance between how accurate you are with what you catch, also depends on how much you miss

  Just being accurate isn't enough - calling afake

  job real might lead to big problems. So, how precise the model is matters just as much as how many it catches.

F. Comparative Analysis

The tests found the **boosted tree** worked better than the forest at spotting fake job ads - using f1-score, precision, recall, and accuracy. Still, the forest might help catch false info on social networks.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

G. System setup to spot fake job ads

The proposed system follows these steps:

1. User sends a job posting or shares key info about the role
2. The system checks the link using **Google Safe Browsing API**
3. Ifthe connection's safe, we pull job info - like title or duties - with web scraping. The site gets checked first before grabbing data from it. Details such as who's hiring show up during extraction. No risky links are used when collecting these posts
4. Data pulled out gets checked by the **Gemini API** - using smart tech tricks to spot signs of scams
5. The system sorts the job into real or fake
6. Info gets saved into a system then shown to you right away

H. Using deep learning methods

To get past the limits of older methods, we used a deep neural network trained on EMSCAD - this dataset holds 18k entries. It's built with:

- Three packed inner layers
- Using NLP to pull out key details from job ads

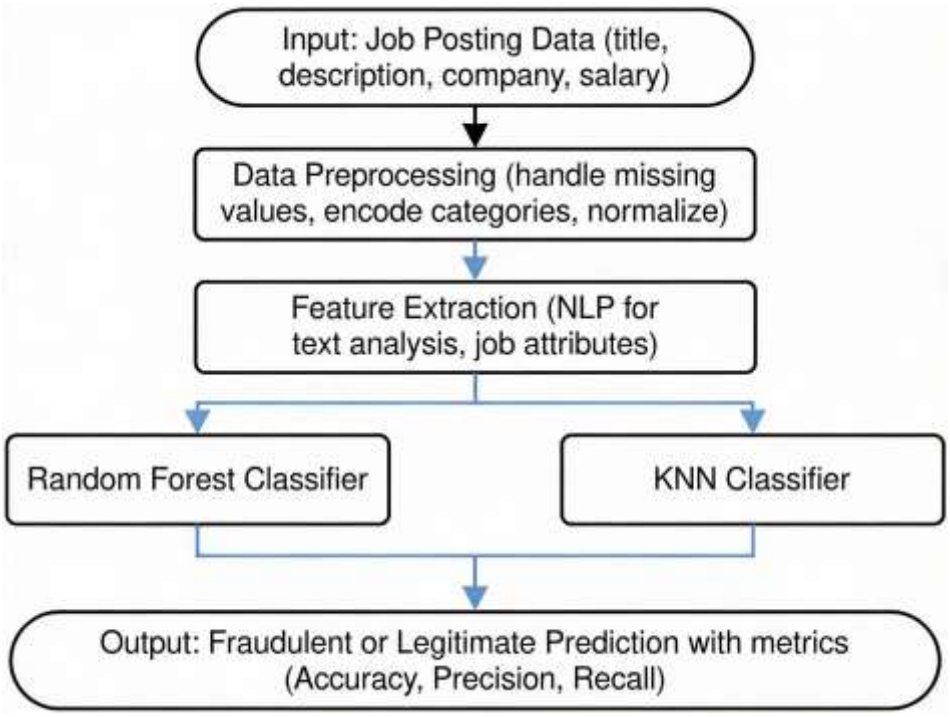
The DNN model hit nearly **98% accuracy** when sorting posts, showing it works better at spotting fake job ads.

I. Putting Together a Random Forest

Random Forest - a method that combines many models - was used here too. Instead of one tree, it builds several while learning from data. Because of this setup, results tend to be more reliable, especially with complex info like job ads. By pulling together different outcomes, it handles messy inputs better than single models.

J. NLP-Based Scam Detection

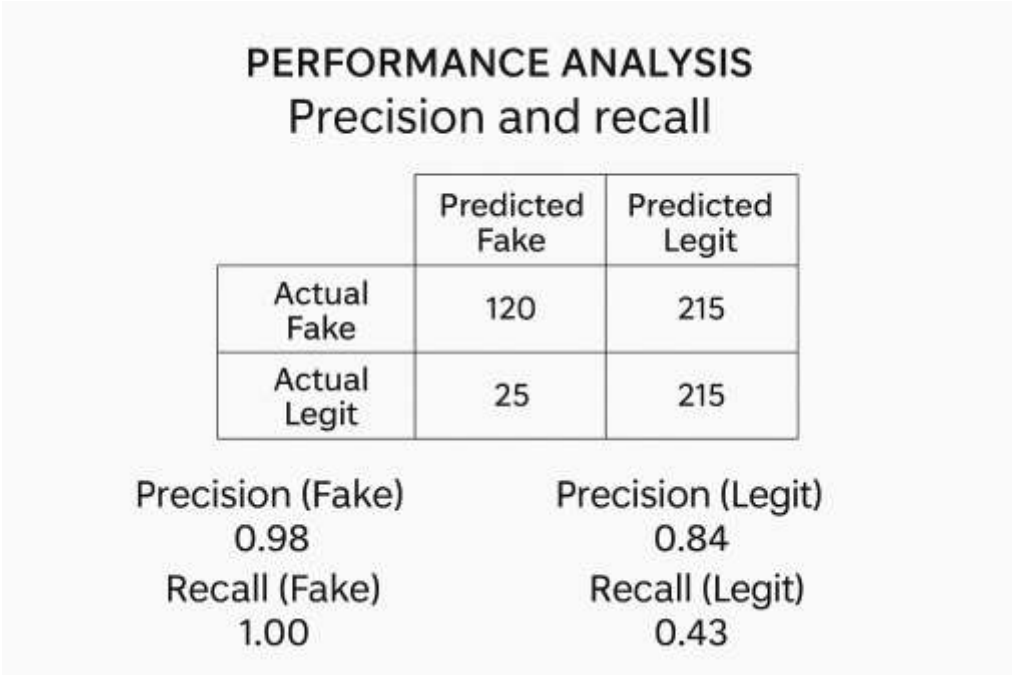
Natural Language



VI. CONCLUSION

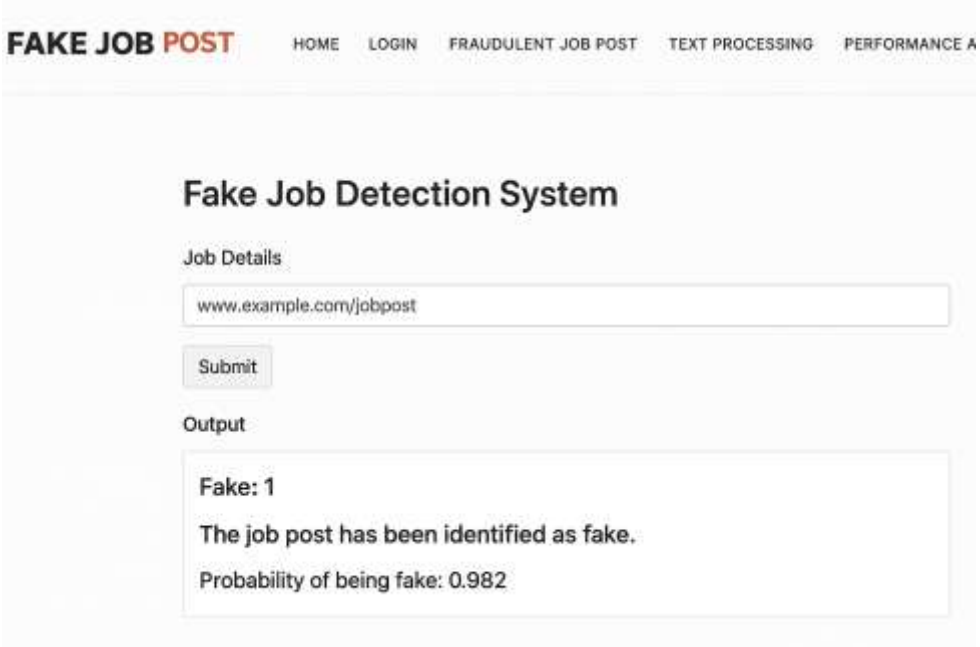
The test outcomes show the new scam-job detector works well telling real ads from fake ones. Not only do the precision-recall curves appear solid, but the confusionmatrix also backs up how reliable themodel is at spotting fakes accurately. It catches most shady listings - so fewer scams slip through unnoticed. All this suggests the method could actually help make job websites safer in real -world use and Output will be:

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal



Acknowledgment
We would like to express our sincere gratitude to our project guide, Mr. Amit Garg, for his continuous support, guidance, and encouragement throughout the development of this project. His insightful feedback and expertise greatly contributed to the successful completion of this work.

We also extend our thanks to our institution and peers for providing the facilities, resources, and motivation that helped shape this project effectively.
Final Output will be:



SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 1 | January 2026 - Peer Reviewed Journal

DECLARATION

We hereby declare that the project titled "Data Intelligence for Identifying Fraudulent Job Opportunities" is an original work carried out by us under the supervision of Mr. Amit Garg. This work has not been submitted to any other institution for the award of any degree or diploma.

REFERENCES

1. *C.S. Anita , P. Nagarajan , G. Aditya Sairam , P. Ganesh , G. Deepakkumar , Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms , ISSN: 2237-0722 Vol. 11 No. 2 (2021) Received: 16.03.2021 – Accepted: 18.04.2021.*
2. *Alandjani , ONLINE FAKE JOB ADVERTISEMENT RECOGNITION AND CLASSIFICATION USING MACHINE LEARNING , Cuadernos de desarrollo aplicados a las TIC, 11(1), 251-267. https://doi.org/10.17993/3ctic.2022.111.251-267*
3. *Shawni Dutta and Prof.Samir Kumar Bandyopadhyay , Fake Job Recruitment Detection Using Machine Learning Approach , International Journal of Engineering Trends and Technology (IJETT) – Volume 68 Issue 4- April 2020.*
4. *Hariharan , Anuseelan , Gokul , Denish ,International Research Journal of Modernization in Engineering Technology and Science , Peer-Reviewed, Open Access, Fully Refereed International Journal) , Volume:07/Issue:03/March-2025 , Impact Factor- 8.187, www.irjmets.com*
5. *Amrutha P, Ms.Kavya H V , "Fake Job Detection," PES Institute of Technology and Management, Shivamogga, Karnataka, India, Zhuzao/Foundry[ISSN:1001-4977] VOLUME 28.*
6. *Lingeshwaran D, Ahamed Khalifa Z , Harish Kumar P B, NRG Sreevani ,Fake Job Post Detection Using Machine Learning: An ADASYN Approach , ISSN 2349-9249 | | May 2025, Volume 12.*
7. *SU Habiba, MK Islam, F Tasnim - 2021 2nd international conference on robotics …, 2021*
8. *FHA Shibly, U Sharma, HMM Naleer Annals of the Romanian Society for Cell Biology 25 (4), 2462-2472, 2021*
9. *Boonthida Chiraratanasopha, Thodsaporn Chay-intr, Current Applied Science and Technology, 10.55003/cast. 2022.06. 22.008 (12 pages)-10.55003/cast. 2022.06. 22.008 (12 pages), 2022*
10. *AH Mohd Hanif, Nurazean Maarop, Norshaliza Kamaruddin, Ganthan Narayana Samy International Journal of Academic Research in Business and Social Sciences 14 (1), 1182-1193, 2024*
11. *K Ramakrishna Reddy, G Indrani, N Pavan Kumar, K Vamshi Krishna, 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 822-827, 2025*
12. *G Srinivas, A Lakshmanarao, S Sushma, M Vamsi Krishna, S Neelimam 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), 1322-1325, 2023*