International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

DETECTING FRAUDULENT JOB POSTINGS USING MACHINE LEARNING MODELS

J.Raghunath[1], Mullah Shaik Afrin[2], Vemuri Bhavana[3], Mandadhi Avanthi[4], Sake Deepak[5], M Naga Vardhan Babu[6], Dasari Mamatha[7], M M Nawaz[8]
raghu.jangam@gmail.com[1], afreenafreen4830@gmail.com[2], vemuribhavana25@gmail.com[3], mandadhiavanthi5@gmail.com[4], sdeepak1451@gmail.com[5], mandatinagavardhan@gmail.com[6], mamathadasari2004@gmail.com[7], nawazmahathar123@gmail.com[8]
DEPARTMENT OF COMPUTER SCIENCE ENGINEERING (Artificial Intelligence)
GATES INSTITUTE OF TECHNOLOGY, Gooty.

ABSTRACT
In modern time, the development in the field of industry and technology has opened a huge opportunity for new and diverse jobs for the job seekers. With the help of the advertisements of these job offers, job seekers find out their options depending on their time, qualification, experience, suitability etc. Recruitment process is now influenced by the power of internet and social media. Since the successful completion of a recruitment process is dependent on its advertisement, the impact of social media over this is tremendous. Social media and advertisements in electronic media have created newer and newer opportunity to share job details. Instead of this, rapid growth of opportunity to share job ads has increased the percentage of fraud job postings which causes harassment to the job seekers. So, people lack in showing interest to new job postings due to preserve security and consistency of their personal, academic and professional information. Thus, the true motive of valid job postings through social and electronic media faces an extremely hard challenge to attain people's belief and reliability. Technologies are around us to make our life easy and developed but not to create unsecured environment for professional life.
If jobs ads can be filtered properly predicting false job ads, this will be a great advancement for recruiting new employees.
Therefore, this project proposed to use different data mining techniques and classification algorithms like K-nearest neighbour, decision tree, support vector machine, naive bayes classifier, random forest classifier, and multi-layer perceptron to predict a job Advertisement 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 classification task. We have used three dense layers for this deep neural network classifier.

INTRODUCTION
The rapid expansion of online job platforms has opened countless employment opportunities, but it has also given rise to an increasing number of fraudulent job postings that exploit job seekers. Scammers create deceptive listings to steal personal information, commit financial fraud, and manipulate unsuspecting individuals. Many job seekers, especially those desperate for employment, fall victim to these scams, leading to significant financial losses and emotional distress. Detecting such fraudulent job postings manually is an overwhelming task due to the sheer volume of listings posted daily. Traditional rule-based methods are often insufficient, as fraudsters continuously adapt their tactics to bypass detection mechanisms. Hence, there is an urgent need for an intelligent, automated system that can efficiently and accurately identify fraudulent job postings in real time.
Machine learning models provide a powerful solution to this problem by analyzing large-scale job posting data, extracting meaningful patterns, and detecting fraudulent characteristics that may not

DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

be immediately apparent to human moderators. By leveraging natural language processing (NLP) techniques, models can identify suspicious keywords, vague job descriptions, and inconsistencies in employer details. Supervised learning algorithms, such as Random Forest, Support Vector Machines (SVM), and deep learning architectures, can be trained on labeled datasets to classify job postings as genuine or fraudulent with high accuracy. Additionally, unsupervised methods, such as anomaly detection, can uncover hidden fraud patterns in evolving scam tactics.The successful implementation of such a model will not only safeguard job seekers from falling victim to employment fraud but also enhance the credibility and security of job portals. By reducing the prevalence of fake job listings, recruitment platforms can provide a safer job search experience, building trust among users and organizations alike. Furthermore, this research contributes to the broader field of cybersecurity and fraud detection, demonstrating the immense potential of AI-driven solutions in combating digital scams. With the growing reliance on online job platforms, it is imperative to integrate machine learning techniques to proactively identify and mitigate fraudulent activities. This project serves as a crucial step toward creating a more secure and reliable job market, ensuring that individuals can pursue career opportunities without fear of deception.

RELATED WORK

Habiba et. al [6] proposed to use different data mining techniques and classification algorithm like KNN, great for this classification task. We have used three dense layers for this deep neural network classifier. decision tree, support vector machine, naïve 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. The trained classifier shows approximately 98% classification accuracy (DNN) to predict a fraudulent job post. Amaar et. al [7] used six machine learning models to analyze whether these job ads are fraudulent or legitimate. Then, we compared all models with both BoW and TF-IDF features to analyze the classifier's overall performance. One of the challenges in this study is our used dataset. The ratio of real and fake job posts samples is unequal, which caused the model over-fitting on majority class data. To overcome this limitation, we used the adaptive synthetic sampling approach (ADASYN), which help to balance the ratio between target classes by generating the number of samples for minority class artificially. We performed two experiments, one with the balanced dataset and the other with the imbalanced data. Through experimental analysis, ETC achieved 99.9% accuracy by using ADASYN as over-sampling ad TF-IDF as feature extraction. Further, this study also performs an in-depth comparative analysis of our proposed approach with state-of-the-art deep learning models and other re-sampling techniques. Mehboob et. al [8] handles the recruitment fraud/scam detection problem. Several important features of organization, job description and type of compensation are proposed and an effective recruitment fraud detection model is constructed using extreme gradient boosting method. It develops an algorithm that extracts required features from job ads and is tested using three examples. The features are further considered for two-step feature selection strategy. The findings show that features of the type of organization are most effective as a stand-alone model. The hybrid composition of selected 13 features demonstrated 97.94% accuracy and outperformed three state-of-the-art baselines. Moreover, the study finds that the most effective indicators are "salary range," "company profile," "organization type," "required education" and "has multiple jobs." The findings highlight the number of research implications and provide new insights for detecting online recruitment fraud. Ranparia et. al [9] minimized the number of such frauds by using Machine Learning to predict the chances of a job being fake so that the candidate can stay alert and take informed decisions, if required. The model will use NLP to analyze the sentiments and pattern in the job posting. The model will be trained as a Sequential Neural Network and using very popular GloVe algorithm. To understand the accuracy in real world, we will use trained model to predict jobs posted on Linked In. Then we worked on improving the model through various methods to

DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

make it robust and realistic. Sudhakar et. al [10] proposed a novel algorithm for classifying phony information and actual news. This study deals with logistic regression, SVM, and novel ensemble approach based on machine learning algorithms. It is divided into sample size values of 620 per group. The experiment uses a dataset of 10,000 records with binary classes (fake news, real news). The result demonstrated that the proposed novel ensemble approach obtains a better accuracy value of 95% and a loss value of 05% compared with other algorithms. Thus, the obtained results prove that the proposed algorithm is an ensemble approach that combines decision tree techniques with AdaBoost by varying parameters and can get a significantly higher accuracy value.

In modern times, advancements in industry and technology have created vast opportunities for job seekers, offering a diverse range of employment options. Online job advertisements help individuals explore career opportunities based on their qualifications, experience, and preferences. The recruitment process has been significantly influenced by the internet and social media, making job advertisements a crucial factor in successful hiring. With the increasing reach of social and electronic media, job postings can now be shared more widely than ever before. However, the rapid expansion of job advertisements has also led to a rise in fraudulent job postings, causing significant challenges for job seekers. Many individuals fall victim to scams, privacy breaches, and misinformation, leading to a loss of trust in new job postings. As a result, job seekers become hesitant to engage with opportunities due to concerns about the security and confidentiality of their personal, academic, and professional information. This lack of trust makes it difficult for legitimate job postings to gain credibility and attract suitable candidates. While technology is designed to enhance convenience and efficiency, it should not contribute to an unsafe professional environment. Filtering out fraudulent job postings is essential to ensuring a secure job-seeking experience. Developing an automated system to detect and predict fake job listings would be a significant advancement in recruitment and Human Resource Management (HRM). Such a system would help job seekers find legitimate opportunities efficiently, reducing time wasted on scams and improving the overall reliability of online job platforms.

PROPOSED SYSTEM

With the increase in online job portals, fraudulent job postings have become a significant concern. These scams not only deceive job seekers but also lead to identity theft, financial loss, and emotional distress. To address this, machine learning techniques are explored to automatically detect fake job advertisements by analyzing patterns in job posting data. This research uses the Employment Scam Aware (EMSCAD) dataset, which contains thousands of labeled job advertisements categorized as real or fake. Each entry includes features such as job title, location, company name, job description, and a label indicating whether it is fraudulent. To prepare the data for model training, several preprocessing steps are performed, including the removal of missing values, encoding of categorical variables, and cleaning of text data. To extract meaningful features from the job descriptions, Term Frequency-Inverse Document Frequency (TF-IDF) is used. This technique transforms textual information into numerical data that machine learning algorithms can process. After feature extraction, various classification models are trained and evaluated, including Naive Bayes, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Multilayer Perceptron (MLP).

The system design follows a sequential workflow, beginning with data preprocessing, followed by feature extraction, model training, and classification. Each model learns patterns associated with fraudulent postings and applies this knowledge to predict new or unseen job ads.

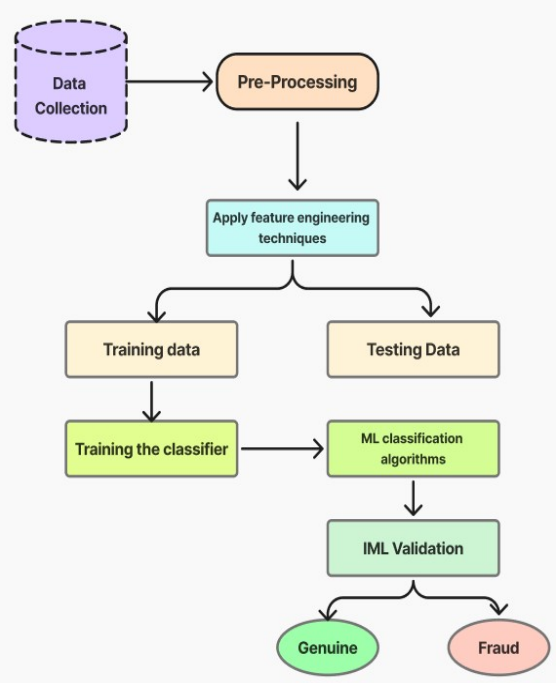
The results demonstrate that machine learning models are capable of distinguishing between real and fake job postings effectively. Among the tested algorithms, neural network-based models showed strong potential due to their ability to capture complex data patterns. This work contributes to creating automated tools that can assist job portals in filtering out fraudulent listings before they reach job seekers.

International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

ADVANTAGES OF PROPOSED SYSTEM

Here are the advantages of the proposed system:
Reduces manual effort by automatically classifying job postings, eliminating the need for human intervention.
Provides instant fraud detection, ensuring job seekers are warned about potential scams before applying.
Can handle a large volume of job postings efficiently, making it suitable for integration with major job portals.
Machine learning models continuously learn from new data, making them effective in detecting emerging scam patterns.
Improves fraud detection precision, reducing the risk of classifying genuine jobs as fake and vice versa.

ARCHITECTURE



DATA FLOW DIAGRAM

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
The data flow diagram (DFD) is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction DFD may be partitioned in to levels that represent increasing information flow and functional detail.
Here’s an example of a Data Flow Diagram (DFD) for the Detecting Fraudulent Job Postings system.

International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

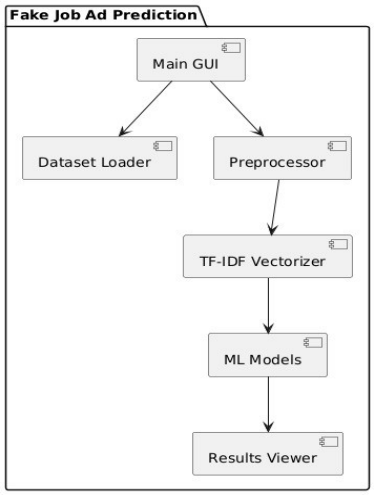


Fig: Data Flow Diagram

RESULTS

Implementation marks a critical phase in the development of any project, where the conceptual or theoretical design is transformed into a practical and functional system. In the context of this project, implementation entails developing a machine learning-based system to detect fraudulent job postings. This phase involves detailed planning, thorough examination of existing methodologies, selection of appropriate tools, and systematic execution. A well-implemented system not only validates the design but also instills confidence in its effectiveness and usability. Additionally, implementation sets the foundation for the evaluation of the system's accuracy and performance in real-world scenarios.

1. Data Collection

This module is the foundational step of the project. It involves gathering a large volume of job-related data from multiple online sources such as job portals, corporate websites, and publicly available datasets like EMSCAD. The dataset includes both legitimate and fraudulent job postings, providing the necessary variability required for training and testing the models. The quality and diversity of the collected data directly impact the effectiveness of the predictive models.

2. Data Preprocessing

The raw data collected is often noisy, inconsistent, and incomplete. Thus, preprocessing is essential to convert it into a usable format. The following steps are performed in this module:

- Removal of duplicate entries and irrelevant data
- Handling missing values using imputation techniques
- Text cleaning (removing special characters, punctuation, stop words, etc.)
- Converting categorical data into numerical formats using encoding techniques
- Tokenizing and vectorizing textual features using techniques like TF-IDF

The output of this module is a clean and structured dataset ready for feature extraction and model training.

3. Feature Selection

This module focuses on identifying the most informative features that contribute significantly to the detection of fraudulent postings. Irrelevant or redundant features are removed to reduce overfitting, improve model performance, and decrease training time. Techniques like correlation analysis, chi-square test, and mutual information are used for selecting the best features. The selected features include job title, company profile, job description, requirements, and employment type.

DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

4. Machine Learning Models

In this phase, various supervised machine learning algorithms are implemented to classify job postings as real or fraudulent. The models used include:

- Support Vector Machine (SVM)
- Decision Tree
- Random Forest
- Naive Bayes
- Multi-Layer Perceptron (MLP)
- K-Nearest Neighbor (KNN)

Each model is trained on the processed dataset and is optimized using suitable hyperparameters. These models learn patterns and characteristics that differentiate legitimate job postings from fraudulent ones.

5. Performance Evaluation

After model training, the final step is to evaluate the performance of each classifier. Standard evaluation metrics such as **Accuracy**, **Precision**, **Recall**, and **F1-Score** are used. These metrics provide a detailed understanding of how well each model is able to detect fraudulent job postings:

- **Accuracy** measures overall correctness.
- **Precision** shows how many predicted fraudulent jobs were actually fraud.
- **Recall** measures how many actual fraudulent jobs were correctly identified.
- **F1-Score** balances precision and recall.

CONCLUSION

Job scam detection has become a great concern all over the world at present. In this project, we have analyzed the impacts of job scam which can be a very prosperous area in research filed creating a lot of challenges to detect fraudulent job posts. We have experimented with EMSCAD dataset which contains real and fake job posts. In this project, we have experimented both machine learning algorithms SVM, KNN, Naive Bayes, Random Forest and a neural network concept called MLP. This work shown the evaluation of machine learning and MLP-based classifiers.

The Detecting Fraudulent Job Postings Using Machine Learning project aims to enhance job seekers' safety by identifying and filtering out fraudulent job postings. With the increasing number of scams in online job portals, a robust fraud detection system is essential to prevent job seekers from falling victim to deceptive employment offers. By leveraging machine learning techniques, the system analyzes job postings based on textual patterns, employer details, salary structures, and other critical features to determine their authenticity. The project successfully integrates data preprocessing, feature extraction, and classification algorithms to detect fraudulent job listings with high accuracy.

Throughout the development and testing phases, various machine learning models were evaluated to determine the most effective approach for fraud detection. The system underwent rigorous testing, including unit testing, integration testing, functional testing, system testing, and acceptance testing, to ensure its reliability and performance. Additionally, white box and black box testing were conducted to verify the system’s internal logic and external behaviour. The results demonstrate that the proposed fraud detection system can effectively flag suspicious job postings while minimizing false positives, making it a practical solution for online job platforms.

Moreover, this project highlights the significance of real-time fraud detection and continuous learning. As scammers evolve their techniques, the system must be regularly updated with new fraudulent patterns and trained on recent datasets to maintain its accuracy. The use of Natural Language Processing (NLP) and AI-driven classification enables the system to adapt and improve over time, ensuring it remains effective against emerging job scams. Additionally, integrating this system with job portals and recruitment websites can provide a proactive defence mechanism, alerting both job seekers and platform administrators about potential threats.

DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

In conclusion, this project successfully demonstrates how machine learning can be applied to enhance job search security and protect users from employment fraud. The implementation of this system can significantly reduce online job scams, improve trust in recruitment platforms, and provide a safer experience for job seekers. Future enhancements may include advanced deep learning models, real-time fraud detection APIs, and integration with government employment verification systems to further strengthen fraud detection capabilities. With continuous development, this project has the potential to revolutionize the way fraudulent job postings are identified and eliminated, creating a safer and more reliable online job market.

FUTURE WORK AND EXTENSIONS

Although the proposed system demonstrates effective performance in detecting fraudulent job postings using various machine learning models, there remains considerable scope for future improvements. One potential direction is the integration of deep learning techniques such as Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and transformer-based models like BERT, which can better capture contextual relationships within job descriptions and improve classification accuracy.

Another promising extension involves the development of a real-time fraud detection system that actively monitors online job portals and flags suspicious postings as they appear. This would significantly enhance the practical usability of the system. Moreover, extending the system to support multilingual datasets would allow its deployment in non-English speaking regions and make it more universally applicable.

Further improvements can be made through advanced feature engineering, utilizing natural language processing techniques such as sentiment analysis, named entity recognition, and semantic similarity. These enhancements would help extract deeper insights from the job description text. Additionally, incorporating a feedback mechanism that allows users to report fraudulent postings could enable the system to learn adaptively and remain effective against evolving fraudulent techniques.

Future developments may also include the creation of a browser extension or mobile application that provides users with real-time alerts about suspicious job ads while browsing. Collaborating with online recruitment platforms to integrate this system directly into their frameworks could ensure broader impact and help maintain a more secure job-seeking environment.

REFERENCES

1. S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, “Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset”, Future Internet 2017, 9, 6; doi:10.3390/fi9010006.
2. B. Alghamdi, F. Alharby, “An Intelligent Model for Online Recruitment Fraud Detection”, of Information Security, 2019, Vol 10, pp. 155¬176, https://doi.org/10.4236/iis.2019.103009.
3. 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.
4. 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.
5. B. Alghamdi and F. Alharby, —An Intelligent Model for Online Recruitment Fraud Detection,” J. Inf. Secur., vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009
6. S. U. Habiba, M. K. Islam and F. Tasnim, "A Comparative Study on Fake Job Post Prediction Using Different Data Mining Techniques," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 543-546, doi: 10.1109/ICREST51555.2021.9331230.

DOI:10.46647/ijetms.2025.v09i02.069 ISSN: 2581-4621

7. Amaar, A., Aljedaani, W., Rustam, F. et al. Detection of Fake Job Postings by Utilizing Machine Learning and Natural Language Processing Approaches. Neural Process Lett 54, 2219–2247 (2022). https://doi.org/10.1007/s11063-021-10727-z
8. Mehboob, A., Malik, M.S.I. Smart Fraud Detection Framework for Job Recruitments. Arab J Sci Eng 46, 3067–3078 (2021). https://doi.org/10.1007/s13369-020-04998-2
9. D. Ranparia, S. Kumari and A. Sahani, "Fake Job Prediction using Sequential Network," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp. 339-343, doi: 10.1109/ICIIS51140.2020.9342738.
10. Sudhakar, M., Kaliyamurthie, K.P. (2023). Efficient Prediction of Fake News Using Novel Ensemble Technique Based on Machine Learning Algorithm. In: Kaiser, M.S., Xie, J., Rathore, V.S. (eds) Information and Communication Technology for Competitive Strategies (ICTCS 2021). Lecture Notes in Networks and Systems, vol 401. Springer, Singapore. https://doi.org/10.1007/978-981-19-0098-3_1