ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Enhanced Online Recruitment Fraud Detection Using Hybrid Neural Networks with CNN-Based Feature Extraction

Mr GHANTASALA VENU GOPAL[1], Maramreddy Navya Sree[2]

#1 Associate Professor in the Department of Computer Science and Engineering at PBR Visvodaya Institute of Technology and Science in Kavali, .

#2M. Techin Computer Science and Engineering at PBR VITS, Kavali, SPSR Nellore, Andhra Pradesh, India

Abstract: The rapid growth of online employment platforms has caused fraudulent job advertisements to become a significant issue in internet recruiting. To solve this problem, we propose a better job categorising tool using modern deep learning algorithms to find false job adverts. Building on the use of transformer-based models like BERT and RoBERTa, we employ a Convolutional Neural Network (CNN2D) to improve feature extraction and maximum classification performance. The CNN2D model precisely and effectively finds fake job ads by use of 2D convolutional layers, hence capturing complex patterns in the dataset. Moreover, the user-friendly interface of the system, which uses the Flask framework, enables simple job posting management and real-time fraud detection. This improvement speeds up and improves the accuracy of the procedure for spotting false employment offers.

Index terms *- Fraudulent Job Postings, Convolutional Neural Network (Cnn2d), Feature Extraction, Real-Time Fraud Detection.*

1. INTRODUCTION

Unpleasant or hazardous side effects resulting from drug use, Adverse Drug Reactions (ADRs) often need medical treatment, dose adjustments, or complete drug cessation. These reactions pose a significant threat to public health systems worldwide since they cause more death, longer hospital stays, and a dramatic rise in healthcare costs. Many negative drug responses (ADRs) go undetected throughout clinical trials and only surface after the medicine has been introduced into the general market, hence early identification and prediction are absolutely vital.

Healthcare quality, geographic area, and sex among other things affect ADR occurrence. Research, for instance, shows that women are more prone to negative drug responses (ADRs) caused by pharmacokinetic and pharmacodynamic differences as well as higher drug doses per body weight. Differences in medical standards and access to treatment among countries could potentially affect ADR reporting and management. Research indicates that a significant portion of ADRs—about 71.6% in affluent countries and 59.6% in poorer ones—can be prevented. Mortality rates linked to ADR are consistent across sites, highlighting the urgent demand for effective forecasting tools.

This work uses Graph Neural Networks (GNNs) and Self-Supervised Learning to explore a deep learning-based approach for forecasting drug-drug interactions and the adverse effects that accompany them in order to solve these challenges. By using structured representations of drugs and their interactions, the model aims to raise the accuracy of ADR identification before drugs reach patients, hence enhancing drug safety and preserving life..

2. LITERATURE SURVEY

In real-world smart systems, machine learning algorithms manage a wide variety of data types. With technological advancements and the widespread use of social media platforms, many recruiters and job seekers are now actively engaging in online job activities. While these innovations enhance accessibility and convenience, they also expose users to potential risks such as privacy invasions and fraudulent schemes. Fraudsters and illegitimate organizations exploit virtual job portals to lure unsuspecting job seekers into scams.

The rise in online job postings, especially in the aftermath of the pandemic, has also seen a proportional increase in fraudulent job advertisements. These deceptive postings not only waste time but also endanger the privacy and personal information of job seekers. Consequently, accurate identification and classification of these fake job ads have become crucial. Machine learning and deep learning techniques have proven to be effective in distinguishing between genuine and fraudulent job postings. Various studies have suggested that classification systems can be significantly improved through effective data preprocessing, cleaning, and feature extraction techniques.

To improve detection accuracy, many approaches utilize a combination of classification models. Comparative studies have explored the performance of different algorithms, such as support vector machines (SVM), decision trees (DT), k-nearest neighbors (KNN), and artificial neural networks (ANN). The use of the Employment Scam Aegean Dataset (EMSCAD) has been prominent in training and evaluating these models. Text preprocessing, tokenization, and natural language processing (NLP) techniques, including TF-IDF vectorization, have been effectively employed to extract meaningful features from job descriptions.

Data purification and preprocessing are essential components of any machine learning pipeline, as they directly impact model performance. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) are used to balance the dataset, especially when dealing with imbalanced class distributions of real versus fake job posts. These methods, when coupled with ensemble classifiers like Random Forest, yield robust results and reduce the risk of overfitting.

Deep learning techniques, particularly artificial neural networks (ANN) and deep neural networks (DNN), have shown promising results in fake job detection. For example, one ANN-based model achieved an accuracy of 91.84%, a recall of 96.02%, and an F-measure of 93.88%, outperforming traditional machine learning methods. DNN architectures with multiple dense layers further enhance classification capabilities, especially when applied to large datasets.

The comparative analysis of multiple classifiers helps identify the most effective model for deployment in real-world systems. This includes evaluating models based on precision, recall, F1-score, and overall accuracy. By integrating machine learning

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

and deep learning into fraud detection systems, it is possible to significantly reduce the number of fake job advertisements and protect job seekers from potential scams

3. METHODOLOGY

i) Proposed Work:

By including a CNN2D-based model for sophisticated feature extraction, the suggested approach improves conventional job categorisation techniques. By means of complicated data patterns, this deep learning technique enables better identification of bogus job ads. A Flask-based web interface that streamlines user interaction supports the system, hence allowing real-time forecasts and quick submission of job posting data. This mix guarantees great classification accuracy and provides a simple interface for administrators to control and track task validity properly..

ii) System Architecture:

The proposed method finds false job ads on recruitment sites using Convolutional Neural Networks (CNN2D). Although transformer-based models such as BERT and RoBERTa have showed potential in spotting false job ads, they can struggle to properly capture complicated data patterns. The system therefore combines the benefits of these transformer models with the enhanced features of CNN2D, which is recognised for its capacity to identify complex patterns in data sets. By using 2D convolutional layers to draw out more subtle characteristics from job ads, the CNN2D model increases the general accuracy and efficiency of the model in differentiating between genuine and fraudulent listings.

The first stage of the system is compiling a thorough dataset comprising both legitimate and fraudulent job postings from different sources. This larger dataset guarantees the system stays current with the most recent job posting trends and helps to offset the constraints of current, obsolete benchmark datasets. The CNN2D model is used to process and extract important aspects after data collection, hence allowing the system to more accurately identify complicated patterns that could suggest fraudulent activity. This hybrid technique uses both the language understanding ability of the transformer models and the feature extraction power of CNN2D in combination with BERT and RoBERTa.

The approach addresses the frequent problem of class imbalance in the dataset by including the Synthetic Minority Oversampling Technique (SMOTE), hence improving the performance of the model. Many SMOTE implementations create fake, fraudulent job advertising to help the system train on a more balanced sample. This enhances the system's capacity to identify fake listings, particularly in situations where bogus jobs are less in number relative to real ads.

The system also has a simple interface built with the Flask framework that offers administrators an easy way to handle job postings. Administrators can post job postings for real-time fraud detection via this interface. The technology provides actionable information by categorising these listings as either real or fake and processing them promptly. Deep learning algorithms coupled with an intuitive interface can quickly and accurately identify false job advertisements, thereby guaranteeing the security of

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

online recruiting platforms.

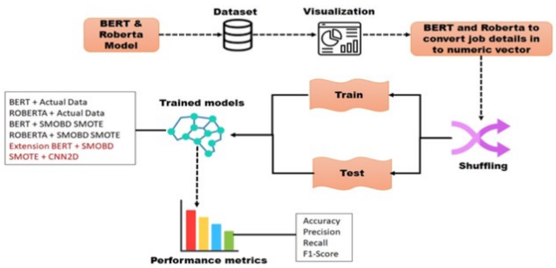

Fig.1. Proposed Architecture

iii) MODULES:

a. **Load BERT & RoBERTa Model**: This module involves importing pre-trained BERT and RoBERTa models from the Hugging Face library. These models serve as powerful text encoders for processing job descriptions and extracting contextual embeddings.

b. **Exploring the Dataset**: In this module, the dataset is examined for structure, content, and potential issues. Key statistics such as class distribution, missing values, and sample entries are analyzed to understand the data.

c. **Visualization**: This module utilizes visualization libraries to create graphs and charts, showcasing insights into the dataset. Visualizations help identify trends, relationships, and anomalies in job postings, enhancing understanding.

d. **BERT and RoBERTa for Vectorization**: Here, BERT and RoBERTa convert job details into numerical vectors. The models tokenize and encode the text, transforming it into dense representations suitable for classification tasks in the subsequent steps.

e. **Shuffling**: This module randomizes the order of data entries to ensure that the training process is unbiased and independent of the input order. Shuffling enhances model robustness and generalization.

f. **Split the data into train & test**: In this module, the dataset is divided into training and testing subsets, typically using an 80-20 split. This separation allows for effective training of the model and evaluation of its performance.

g. **Model generation:** Model building – BERT + Actual Data, ROBERTA + Actual Data, BERT + SMOBD SMOTE, ROBERTA + SMOBD SMOTE, Extension BERT + SMOBD SMOTE + CNN2D. Performance evaluation metrics for each algorithm is calculated.

h. **Admin login**: In this module, admin can login into the application.

i. **Predict Fraud Job**: In this module user can upload the input data.

j. **Logout**: User can logout after the completion of all activities.

4. EXPERIMENTAL RESULTS

Using the Basic Neural Network approach with features from BERT and ROBERTA, the model was able to classify job postings into real or fake, with a reasonable accuracy. However, when CNN2D was applied, the accuracy significantly improved. The CNN2D layers helped the model capture more complex patterns in the data, resulting in better feature optimization and classification performance.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

When BERT and ROBERTA were combined with the SMOTE SMOBD algorithm (for balancing the dataset), they performed exceptionally well in detecting fake job postings, with BERT showing higher accuracy compared to ROBERTA. The CNN2D extension further boosted accuracy, as it could better extract and process features, making it more effective than a basic neural network.

All datasets can be downloaded from below URL

https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction
https://www.kaggle.com/datasets/promptcloud/indeed-job-posting-dataset

https://www.kaggle.com/datasets/zusmani/pakistans-job-market

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

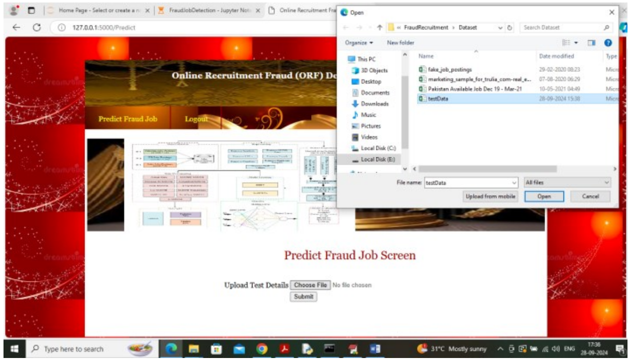

Fig.2. upload test file

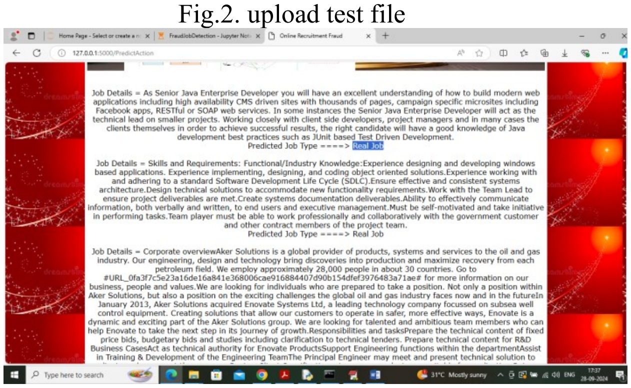

Fig.3. dataset analysis

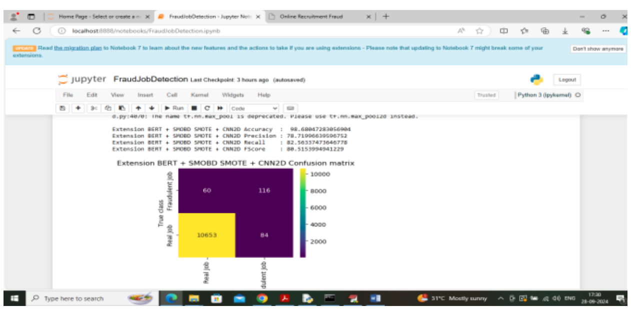

Fig.5. In above screen extension got 98.68% accuracy which is higher than all propose algorithms and in confusion matrix can see both real and fake jobs predicted

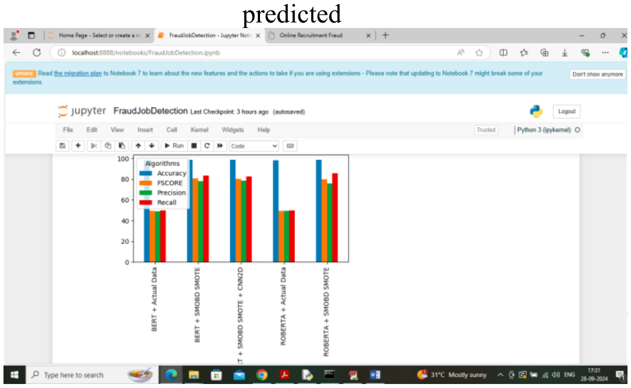

Fig.5. algorithm performance graph

5. CONCLUSION

The expansion idea of classifying job ads with CNN2D combined with BERT and ROBERTA has shown to greatly improve the performance of the model. Although the fundamental neural network method offered a strong basis, the advent of CNN2D for greater feature extraction and optimisation, hence improving accuracy. The model could efficiently handle data imbalance and find false job ads more precisely by merging BERT and ROBERTA with the SMOTE SMOBD algorithm. All things considered, this extension idea outperforms conventional methods in both accuracy and feature extraction capacity by providing a strong solution for precisely spotting bogus job ads.

6. FUTURE SCOPE

Using ensemble methods and advanced feature extraction algorithms, the study aims to find online recruiting fraud. Combining attention mechanisms with recurrent neural networks (RNNs) could enable the model to more effectively gather contextual information in job advertisements. Transfer learning from pre-trained models can maximise performance on sma.

REFERENCES

[1] G. Othman Alandjani, ‘‘Online fake job advertisement recognition and classification using machine learning,’’ 3C TIC, Cuadernos de Desarrollo Aplicados a las TIC, vol. 11, no. 1, pp. 251–267, Jun. 2022.

[2] A.Adhikari, A.Ram, R.Tang, and J.Lin, ‘‘DocBERT: BERT for document classification,’’ 2019, arXiv:1904.08398.

[3] I. M. Nasser, A. H. Alzaanin, and A. Y. Maghari, ''Online recruitment fraud detection using ANN,'' in Proc. Palestinian Int. Conf. Inf. Commun. Technol. (PICICT), Sep. 2021, pp. 13–17.

[4] C.Lokku, ''Classification of genuinity in job posting using machine learn ing,'' Int. J. Res. Appl. Sci. Eng. Technol., vol. 9, no. 12, pp. 1569–1575, Dec. 2021.

[5] S. U. Habiba, Md. K. Islam, and F. Tasnim, ''A comparative study on fake job post prediction using different data mining techniques,'' in Proc. 2nd Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST), Dhaka, Bangladesh, Jan. 2021, pp. 543–546.

[6] Report Cyber. Accessed: Jun. 25, 2022. [Online]. Available: https://www.actionfraud.police.uk/

[7] S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, ''Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset,'' Future Internet, vol. 9, no. 1, p. 6, Mar. 2017.

[8] S. Dutta and S. K. Bandyopadhyay, ''Fake job recruitment detection using machine learning approach,'' Int. J. Eng. Trends Technol., vol. 68, no. 4, pp. 48–53, Apr. 2020.

[9] B. Alghamdi and F. Alharby, ''An intelligent model for online recruitment fraud detection,'' J. Inf. Secur., vol. 10, no. 3, pp. 155–176, 2019.

[10] S. Lal, R. Jiaswal, N. Sardana, A. Verma, A. Kaur, and R. Mourya, ''ORFDetector: Ensemble learning based online recruitment fraud detec tion,'' in Proc. 12th Int. Conf. Contemp. Comput. (IC3), Noida, India, Aug. 2019, pp. 1–5.

[11] G. Kovács, ''An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets,'' Appl. Soft Comput., vol. 83, Oct. 2019, Art. no. 105662.

[12] S. Gazzah and N. E. B. Amara, ''New oversampling approaches based on polynomial fitting for imbalanced data sets,'' in Proc. 8th IAPR Int. Workshop Document Anal. Syst., Nara, Japan, Sep. 2008, pp. 677–684.

[13] O. Nindyati and I. G. Bagus Baskara Nugraha, ''Detecting scam in online job vacancy using behavioral features extraction,'' in Proc. Int. Conf. ICT Smart Soc. (ICISS), vol. 7, Bandung, Indonesia, Nov. 2019, pp. 1–4.

[14] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, ''Handling imbalanced datasets: A review,'' GESTS Int. Trans. Comput. Sci. Eng., vol. 30, no. 1, pp. 25–36, 2006.

[15] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, ''Toward credible evaluation of anomaly-based intrusion-detection methods,'' IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 5, pp. 516–524, Sep. 2010.

[16] Y.-H. Liu and Y.-T. Chen, ''Total margin based adaptive fuzzy support vector machines for multiview face recognition,'' in Proc. IEEE Int. Conf. Syst., Man Cybern.,

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Waikoloa, HI, USA, Oct. 2005, pp. 1704–1711.

Author's Profile

Mr GHANTASALA VENU GOPAL is an Associate Professor in the Department of Computer Science and Engineering at PBR Visvodaya Institute of Technology and Science in Kavali, Andhra Pradesh, India. With 22years of academic experience.

His qualification AMIE(COMPUTER ENGINEERING), M.TECH(CSE), Ph.D in Computer science from Rayalaseema University, Kurnool

I am Maramreddy Navya Sree, currently pursuing a M. Techin Computer Science and Engineering at PBR VITS, Kavali, SPSR Nellore, Andhra Pradesh, India. My areas of interest include Testing, SAP and Data Analytics . I have earned certifications such 'Testing' from Institute and present SAP course pursuing. I have certificate on International Conference on Recent Advances in Computing and Electronics.