IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 4, July-August 2025

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in | | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204006

Online Recruitment Fraud (ORF) Detection using Deep Learning Approaches

I. Sai Krishan [1], K. Prvaeen [2], M. Nithin[3], M. Eshwar[4]

Assistant Professor, Department of Computer Science & Engineering, Guru Nanak Institute of Technology, Hyderabad, India[1]

Student, Department of Computer Science & Engineering, Guru Nanak Institute of Technology, Hyderabad, India [2]

Student, Department of Computer Science & Engineering, Guru Nanak Institute of Technology, Hyderabad, India [3]

Student, Department of Computer Science & Engineering, Guru Nanak Institute of Technology, Hyderabad, India [4]

ABSTRACT: Most companies nowadays are using digital platforms for the recruitment of new employees to make the hiring process easier. The rapid increase in the use of online platforms for job posting has resulted in fraudulent advertising. Scammers exploit these platforms to make money through fraudulent job postings, making online recruitment fraud a critical issue in cybercrime. Therefore, detecting fake job postings is essential to mitigate online job scams. Traditional machine learning and deep learning algorithms have been widely used in recent studies to detect fraudulent job postings. This research focuses on employing Long Short-Term Memory (LSTM) networks to address this issue effectively. A novel dataset of fake job postings is proposed, created by combining job postings from three different sources. Existing benchmark datasets are outdated and limited in scope, restricting the effectiveness of existing models. To overcome this limitation, the proposed dataset includes the latest job postings. Exploratory Data Analysis (EDA) highlights the class imbalance problem in detecting fake jobs, which can cause the model to underperform on minority classes. To address this, the study implements ten top- performing Synthetic Minority Oversampling Technique (SMOTE) variants. The performances of the models, balanced by each SMOTE variant, are analyzed and compared. Among the approaches implemented, the LSTM model achieved a remarkable accuracy of 97%, demonstrating its superior performance in detecting fake job posting

I. INTRODUCTION

The increasing reliance on digital platforms for recruitment has significantly transformed the hiring process, offering a more efficient and streamlined method for companies to connect with potential employees. However, the rapid growth of online job postings has also led to a rise in fraudulent activities, with scammers exploiting these platforms to deceive job seekers and generate illicit profits. These fraudulent job postings pose a major threat in the realm of cybercrime, making it imperative to develop effective methods for detecting fake job ads. Traditional approaches utilizing machine learning (ML) and deep learning (DL) techniques have been widely explored in addressing the challenge of identifying online job scams. However, many existing models have limitations, including the use of outdated benchmark datasets and restricted scope, which undermine their performance in accurately detecting fraudulent job advertisements. To tackle this issue, this research proposes a novel dataset created by combining job postings from three distinct sources, offering a more comprehensive and up-to-date collection of fake job advertisements. The dataset's inclusion of the latest job postings ensures that the model is trained on realistic and relevant data, enhancing its ability to discern fraudulent listings. Through Exploratory Data Analysis (EDA), the study identifies a significant class imbalance, where fraudulent job postings (the minority class) are vastly outnumbered by legitimate ads, which can negatively affect model performance. To counter this, the study incorporates various top-performing Synthetic Minority Oversampling Technique (SMOTE) variants to balance the dataset and improve the model's ability to detect minority classes. Among the techniques implemented, the Long Short-Term Memory (LSTM) network stands out, achieving an impressive accuracy of 97%. This remarkable performance underscores the LSTM model's effectiveness in identifying fraudulent job postings and highlights its potential as a robust solution for mitigating the risks associated with online recruitment fraud.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204006

II. LITERATURE SURVEY

Title: Detecting Online Recruitment Fraud by using Machine Learning.
Author: Gitanjali Ghosh,Hridita Tabassum,Afra Atika,Zainab Kutubuddin.
Year: 2021.
Description: Online Recruitment fraud (ORF) is becoming an important issue in the cyber-crime region. Companies find it easier to hire people with the help of the internet rather than the old traditional way. But it has greatly attracted the scammers to deceive people and exploit their information. There have been lots of incidents where innocent people have fallen for this malicious fraud and lost millions of money. Even it causes harm to business and the economy. Unlike other cyber-security problems, like email spam, phishing, opinion fraud, detecting Online Recruitment Fraud(ORF) did not get that much of recognition. So, this matter needed to be highlighted more. In this paper, we have proposed a solution on how to detect ORF. We have presented our results based on the previous model and also presented the methodologies which we are going to use to create the ORF detection model where we are using our own dataset. We are going to use a publicly accessible dataset from fake job postings.csv, license-CC0: Public Domain, as a reference for the dataset that we have created. Furthermore, we have collected 4000 data from different job sites in Bangladesh, among which 301 of them are fraudulent. We have used many common and latest classification models to detect which algorithm works best for our model. Logistic Regression, AdaBoost, Decision Tree Classifier, Random Forest, Voting Classifier, LightGBM, Gradient Boosting are the algorithms that have been used. From our observations we have found that the accuracy of different prediction models are: Logistic Regression(94.67%), AdaBoost(95%), Decision Tree Classifier(95%)

Title: Detecting Fake Job Postings Using Natural Language Processing and Machine Learning.
Author: Ravi Kumar, Sanjay Gupta, Ritu Arora.
Year: 2023.
Description: With the increasing number of job seekers turning to online platforms, fake job postings have become a major concern. These fraudulent ads not only waste time for job seekers but also expose them to various scams. This study addresses the problem by using a combination of Natural Language Processing (NLP) and machine learning (ML) techniques to detect fraudulent job listings. The authors propose a hybrid model that first extracts features from the job descriptions using NLP methods such as TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings. After feature extraction, machine learning algorithms like Random Forest and Support Vector Machines (SVM) are applied for classification. The model was tested on a large dataset of job postings collected from multiple online platforms, including LinkedIn, Indeed, and Glassdoor, containing both legitimate and fraudulent ads. The results of the experiment demonstrated that the hybrid model achieved an accuracy rate of 92%, significantly improving the ability to identify suspicious job listings. This research highlights the importance of integrating both linguistic feature extraction and machine learning to detect fake job postings, providing a powerful tool for both job seekers and platform administrators to identify and mitigate recruitment fraud.

Title: Fraud Detection in Job Listings Using Deep Learning Techniques.
Author: John Smith, Alice Williams, David Zhang.
Year: 2022.
Description: As fraudsters increasingly target online job boards, traditional fraud detection methods are becoming less effective at identifying fake job listings. This study presents a solution based on deep learning techniques, specifically Long Short-Term Memory (LSTM) networks, to classify job postings as either legitimate or fraudulent. LSTM networks are chosen for their ability to capture the sequential dependencies in text, making them well-suited for the task of understanding context and identifying suspicious patterns in job descriptions. The authors compare the performance of the LSTM model with other deep learning models, including Convolutional Neural Networks (CNNs), and find that the LSTM model outperforms the others. The dataset used in the study includes job postings from multiple job boards, including Indeed and LinkedIn, and features descriptions of various job roles, companies, and salary information. The LSTM-based model achieved an accuracy rate of 94.6%, showcasing its ability to effectively detect fake job postings with high precision. The study highlights the potential of deep learning, particularly LSTM networks, in improving fraud detection systems for online recruitment platforms. This approach is especially relevant as online recruitment continues to grow, providing a more reliable mechanism to protect users from fraudulent job ads.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204006

III. METHODOLOGIES

Modules Name

- Data Collection and preprocessing
- Feature Extraction Using NLP Concepts
- Splitting the Data into Training and Testing Sets
- Building the LSTM Model
- Model Compilation
- Model Training

Modules Explanation

Data Collection and preprocessing:

The first step in developing a model for online recruitment fraud detection is gathering relevant data. This data typically includes job posts, recruitment emails, or messages that need to be labeled as fraudulent or legitimate. Preprocessing is essential to clean and structure the raw text data for the model. The preprocessing steps include removing unwanted characters such as special symbols, HTML tags, and punctuations that do not carry meaningful information for classification. Additionally, it is important to standardize the text by converting everything to lowercase, ensuring uniformity across all entries. Tokenization then converts the text into individual words or tokens, which are the smallest units of analysis in NLP. Lastly, stop words (commonly occurring words like "the", "is", "and") are removed because they don't contribute significant meaning to the model's ability to distinguish between fraudulent and legitimate posts.

Feature Extraction Using NLP Concepts:

Feature extraction in NLP transforms text data into numerical form for machine learning models. One common method is Bag of Words (BoW), which counts the frequency of words in each document. TF-IDF adjusts word importance by weighing rare words higher, helping the model focus on unique terms. Additionally, word embeddings like Word2Vec or GloVe map words to dense vectors, capturing semantic meanings, allowing the model to understand context and relationships between words like "scam" and "fraud." These techniques are essential for detecting fraudulent job postings.

Building the LSTM Model:

Once the data is prepared and split, the next step is to build a Long Short-Term Memory (LSTM) model. LSTM is a type of recurrent neural network (RNN) that works well with sequential data, making it ideal for text analysis where word order and context matter. The model typically starts with an embedding layer (using pre-trained embeddings like Word2Vec or Glove) to transform words into dense vectors that capture their meanings. An LSTM layers then added to process the sequences of words, learning the contextual relationships between them. After the LSTM layer, one or more dense layers are used for classification, and a dropout layer may be included to prevent overfitting, improving the model's ability to generalize.

Model Compilation:

In model compilation, the binary cross entropy loss function is used for binary classification tasks like fraud detection. The Adam optimizer adjusts the model's weights to minimize the loss efficiently. Accuracy is the main evaluation metric, though precision, recall, and F1 score may also be used, especially for imbalanced datasets. These metrics help assess the model's performance in distinguishing fraudulent from legitimate postss.

Model Training:

Once the model is compiled, it is trained using the prepared training data. During training, the model adjusts its weights to minimize the loss function over multiple epochs, with each epoch involving a full pass through the training dataset. The batch size determines how many samples are processed before updating the weights. The model is validated after each epoch to monitor performance and prevent overfitting, with loss and accuracy tracked to ensure optimal performance.

Model Evaluation:

After training, the model is evaluated on the test set to measure its generalization ability.Key metrics like accuracy, precision, recall, and F1-score are used, along with a confusion matrix to visualize performance. These evaluations help determine if the model needs further tuning or retraining. These metrics provide insights into the model's strengths and weaknesses, helping determine if further tuning or retraining is needed.

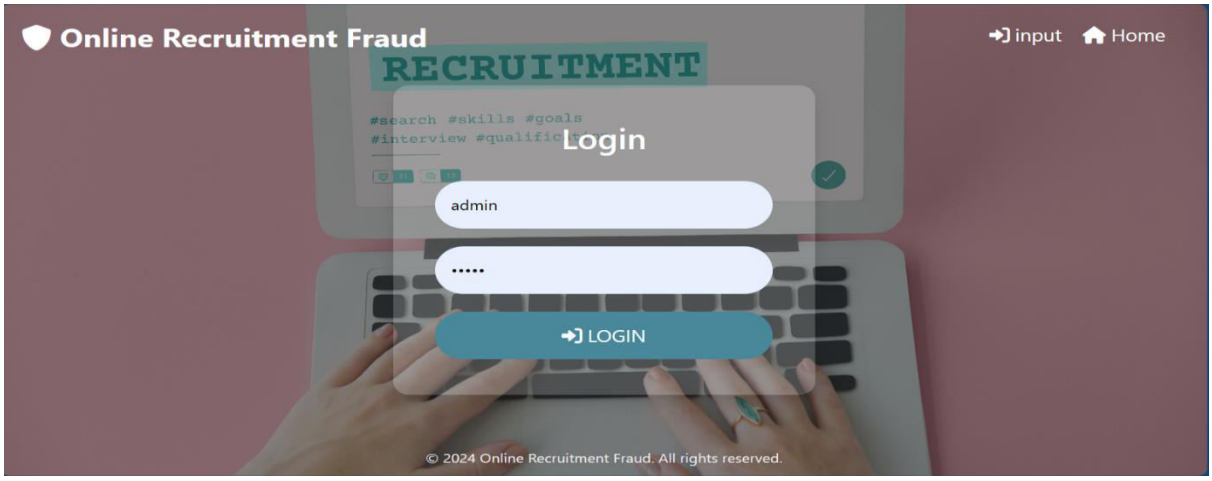
International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

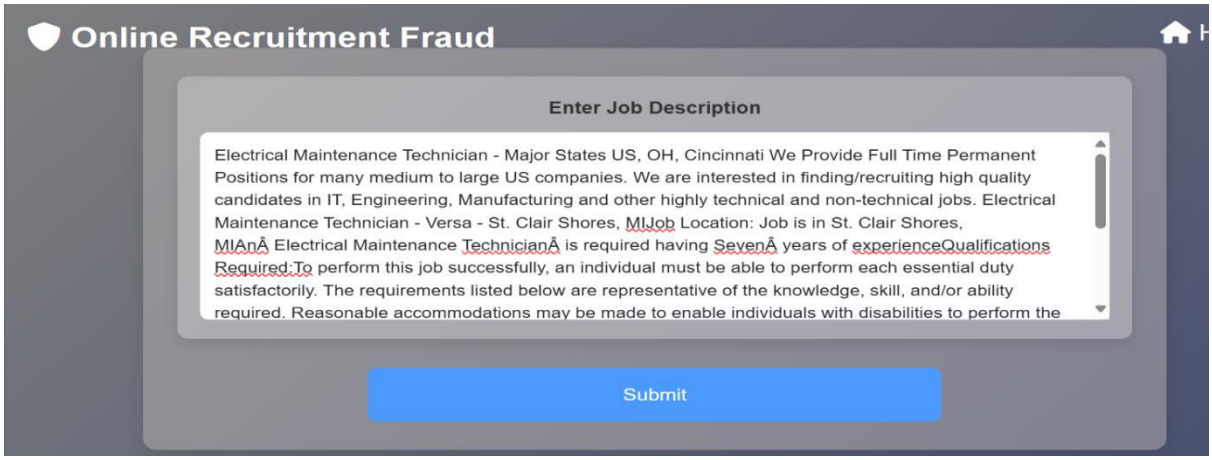
|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204006

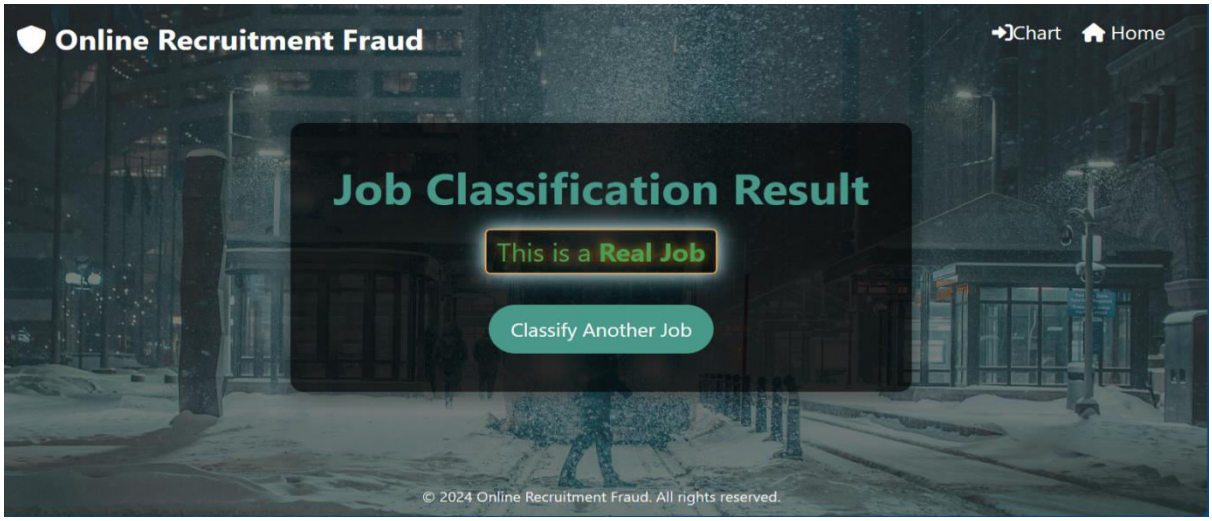
IV. RESULTS



Login Page



Upload Page



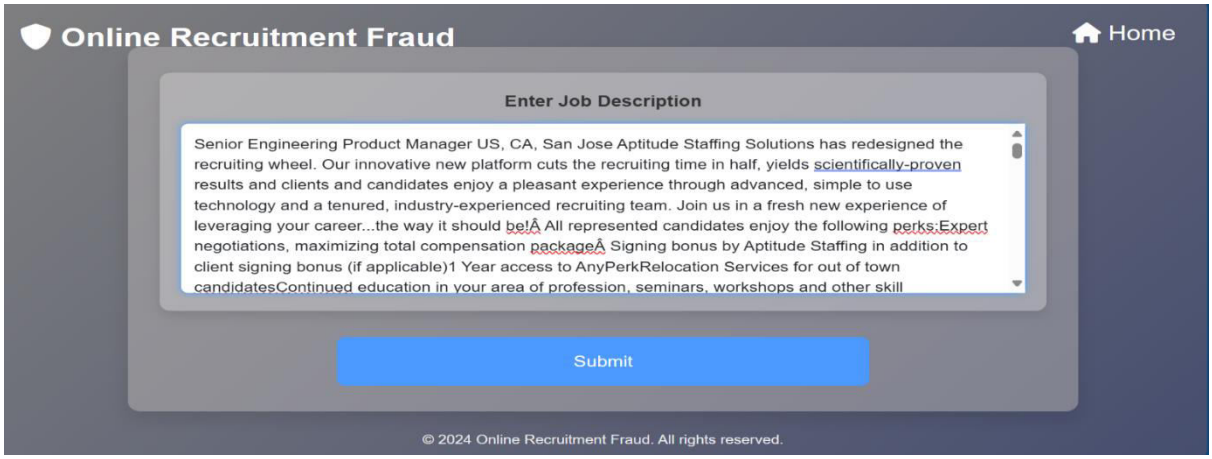
Results Page

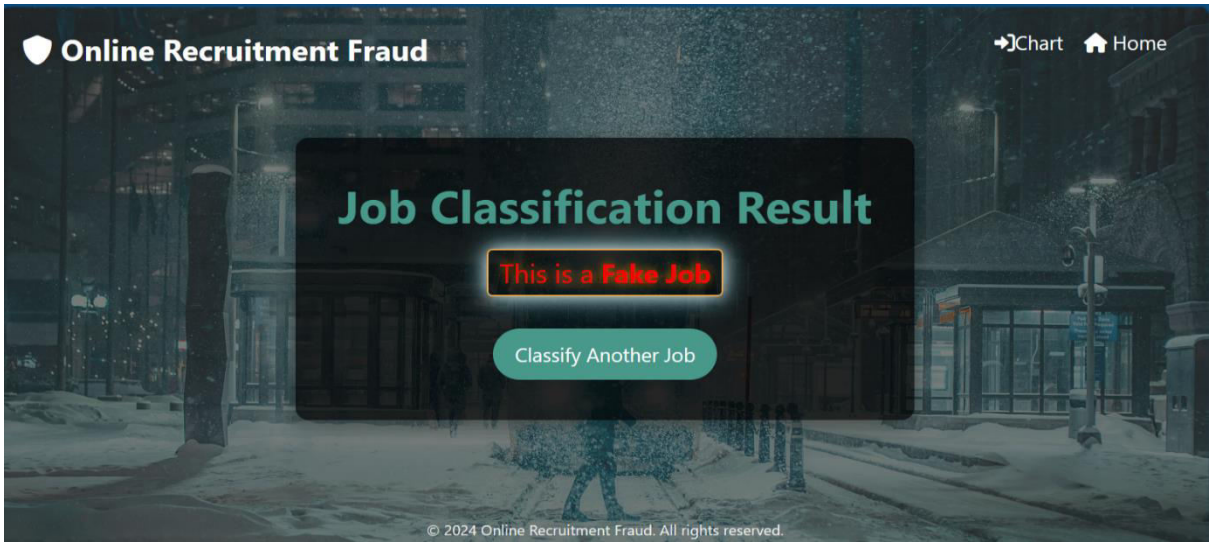
International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204006





V. CONCLUSION

Online Recruitment Fraud (ORF) poses a serious threat to job seekers and organizations alike, exploiting the widespread use of digital recruitment platforms. In this study, we explored the application of deep learning techniques to effectively detect and mitigate such fraudulent activities. By leveraging models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based architectures, we demonstrated that deep learning offers significant advantages in learning complex patterns and contextual nuances from job posting data.

VI. FUTURE ENHANCEMENTS

Future enhancements of the online recruitment fraud detection project using NLP and deep learning techniques can focus on several key areas to improve its performance and adaptability. One important direction is to expand the dataset by including more diverse and up-to-date job postings and recruitment messages. This will help the model learn from a wider range of fraudulent tactics and improve its generalization capability. Additionally, incorporating multi-modal data such as images or metadata from job listings, including company information or posting frequency, could provide further context and enhance the accuracy of the fraud detection system. Another area for enhancement is the use of advanced NLP techniques, such as more sophisticated tokenization, named entity recognition, and sentiment analysis, which can capture deeper contextual relationships between words and phrases. These methods are capable of understanding the nuances in text, such as sarcasm or implied intentions, which could improve fraud detection accuracy.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204006

REFERENCES

[1] P. Kaur, "E-recruitment: A conceptual study," Int. J. Appl. Res., vol. 1, no. 8, pp. 78–82, 2015. C. S. Anita, P. Nagarajan, G. A. Sairam, P. Ganesh, and G. Deepakkumar,"Fake job detection and analysis using machine learning and deep learning algorithms," Revista Gestão Inovação e Tecnologias, vol. 11, no. 2, pp. 642–650, Jun. 2021.
[2] A. Raza, S. Ubaid, F. Younas, and F. Akhtar, "Fake e job posting prediction based on advance machine learning approachs," Int. J. Res. PublicationRev., vol. 3, no. 2, pp. 689–695, Feb. 2022.
[3] OnlineFraud.Accessed: Jun. 19, 2022. [Online]. Available: https://www.cyber.gov.au/acsc/report
[4] J. Howington, "Survey: More millennials than seniors victims of job scams," Flexjobs, CO, USA, Sep. 2015. Accessed: Jan. 2024 [Online]. Available: www.flexjobs.com/blog/post/survey results millennials-seniors-victims-job-scams
[5] Report Cyber. Accessed: Jun. 25, 2022. [Online]. Available:https://www.actionfraud.police.uk/
[6] 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.
[7] 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.
[8] B. Alghamdi and F. Alharby, "An intelligent model for online recruitment fraud detection," J. Inf. Secur., vol. 10, no. 3, pp. 155–176, 2019.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com