Proceedings of the 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV-2025)
IEEE Xplore Part Number: CFP25ONG-ART; ISBN: 979-8-3315-1175-3

2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) | 979-8-3315-1175-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICICV64824.2025.11085449

AI-Powered Detection of Fraudulent Job Listings using DistilBERT and XGBoost

Mr Keerthi Shetty
Professor, Department of MCA
N.M.A.M Institute of Technology
Nitte (Deemed to be University)
Nitte, Karnataka, India
keerthi.shetty@nitte.edu.in

Saiprasad K Shetty
Postgraduate Student, Department of MCA
N.M.A.M Institute of Technology
Nitte (Deemed to be University)
Nitte, Karnataka, India
saiprasadshetty85@gmail.com

Sanjana B
Postgraduate Student, Department of MCA
N.M.A.M Institute of Technology
Nitte (Deemed to be University)
Nitte, Karnataka, India
bsanjanamadival@gmail.com

Dr. Anantha Murthy
Professor, Department of MCA
N.M.A.M Institute of Technology
Nitte (Deemed to be University)
Nitte, Karnataka, India
anantham2004@nitte.edu.in

Ms. Harshitha G M
Professor, Department of MCA
N.M.A.M Institute of Technology
Nitte (Deemed to be University)
Nitte, Karnataka, India
harshitha.gm@nitte.edu.in

Ms. Prathwini
Professor, Department of MCA
N.M.A.M Institute of Technology
Nitte (Deemed to be University)
Nitte, Karnataka, India
prathwini.devadiga@nitte.edu.in

***Abstract*—Fake job ads pose significant danger to job hunters on the Internet because they can steal personal data or money. This research demonstrate a model which is AI-driven using DistilBERT for text understanding and XGBoost for classification which successfully identifies the fraudulent job postings. NLP is employed to understand the job descriptions in addition to the characteristics such as the type of telecommuting job. The application of SMOTE(Synthetic Minority Oversampling Technique) to the data in combination with the advanced machine learning methods led to the successful model outcome, producing high accuracy and a reliable outcome. The research concludes that the system, which is able to identify fake job advertisements, thus reducing the number of fraud occurrences on the job market, is a feasible solution to online job market safety.**

***Index Terms*—SMOTE (Synthetic Minority Oversampling Technique), DistilBERT, XGBoost, Natural Language Processing (NLP)**

I. Overview

Job-related scams has increased online. These methods are not just about a job, they target the job but the final result is financial and emotional suffering for the victims. Trusted ways of identifying fake job posts like checking a person's professional skill that was acquired and human interaction which however, does not help secure the process of reading, can be used. Those who review the scams find it difficult to as they can miss the fraud patterns due to being overwhelmed by the volume of job advertisements on online platforms.

One of the issues or a research problem that is addressed in this article is building an automated AI-based tool to detect fake job ads using Natural Language Processing and machine learning approaches. This tool just cannot be as robust and scalable as the robotic framework, one that processes only structured data and may even detect specific types of data in a different way. This paper has identified potential areas for traditional recursive methods in the classification of machine-based learning and also strict data filter developers. The model's aptitude to put in-order large databases, distinguish cheater patterns, and propose detection that portal time have made it a deserving choice for inclusion in skill find platforms. The ultimate goal of the presented approach is to protect job seekers by making the trust level higher and securing the whole of the online job markets. Machine learning methods have also been studied extensively in other fraud offenses, i.e., credit card fraud, to which data-driven solutions are a solution [9]. .

II. Literature Survey

Several studies have attempted job fraud detection using different approaches:

A. Machine Learning Models

The past few algorithms have used Support Vector Machine (SVM), Random Forest and Logistic Regression, etc. These models promptly acquire linear and non-linear associations in the formatted information, though their efficiency can decrease if the datasets are over-scaled. Earlier research has tried online recruitment scam detection with diverse ML models which aimed at early identification to protect job seekers [1].

B. Transformer Models

Ongoing research shows BERT and DistilBERT-based models to be effective for relevant text classification tasks. Recent research has highlighted the way pre-trained models such as DistilBERT uses deep contextual representations to achieve greater accuracy in NLP tasks [12]. The proposed self-attention mechanism which is integrated into these transformer models has the capability to examine job descriptions and company profiles and detect the out-of-the-ordinary patterns

that lead to detection inaccuracy at a very high level of accuracy.

C. Feature Engineering Techniques

Researchers discovered the patterns that lead to fraud incidents among job attributes, such as telecommuting, company logo and unstructured text like job descriptions. Feature engineering in the form of NLP embeddings and word vectors as well as tools like word representation learning can also be helpful in attempt to understand the structure and meaning of textual data using deep data structure analysis. These techniques are based on fundamental NLP principles, such as semantic embeddings and context analysis [13].

D. Imbalance Handling

In a large number of instances where fraud has occurred in the job sector, genuine job listings indeed outnumber the ones that are fraudulent. Statistical methods such as the generation of synthetic samples and the even distribution of data through the minority class are often used to avoid the problem of the imbalance issue and the subsequent learning by majority classes only.

III. Methodology

A. Dataset Collection & Preprocessing

1) Source: The dataset is made up of job attributes organized as rows in a table, and textual descriptions that are taken from job portals all over the Internet.

2) Cleaning: Data Preprocessing removes missing values and duplicates, as well as getting rid of irrelevant features. Textual data is cleaned using normalization techniques, including lower casing, removing punctuation, and tokenization.

3) Feature Engineering: It covers the use of DistilBERT for extracting job descriptions and thus assembling feature words. Besides this, one-hot encoding is used for transforming the categorical variable and numerical features are transformed to a standard scale.

B. Handling Class Imbalance

Class imbalance remains a core challenge in fraud detection, severely impacting model generalization [17]. The dataset exhibits a significant class imbalance, with genuine job postings vastly outnumbering fraudulent ones. Balancing imbalanced data is also a problem in classification, which has been a well-established problem as one of the primary challenges in supervised learning domains [12] that need methods like SMOTE in order to have effective and balanced training outcomes [11]. Techniques like SMOTE have been shown in literature to perform effectively in oversampling imbalanced datasets, particularly for classification [2].

1) SMOTE (Synthetic Minority Over-sampling Technique): This technique is applied to create a fair representation of both the majority and minority classes making sure there is a balanced dataset for model training.

C. Model Selection and Training

1) DistilBERT: Transformer-based models like BERT have shown state-of-the-art performance in various NLP tasks [15]. DistilBERT, the less computationally costly version of BERT that maintains language understanding with fewer parameters, proved to be efficient for text classification [3].It has been leveraged for the purpose of the pre-processing of the job descriptions in terms of their textual features, hence reflecting different implied contexts in terms of the language. High frequency of word occurrences or linkages between words to others may affect the calculations adding to or reducing their weights.

2) XGBoost: Theoretically, boosting algorithms are derived from the research by Freund and Schapire (1997) [14] and extended to work on scalable models like XGBoost. XGBoost is among the most efficient and scalable gradient boosting classifiers for classification tasks [8]. The embedding of words of a certain job opening is obtained when XGBoost is applied to the job embeddings. Hereby, the model is provided with a series of features and their labels. The model will utilize them during the training to learn to assign the correct label to each of the features. Though XGBoost remains a good classifier, boosting algorithms like LightGBM also offer competitive performance with better speed and memory usage [4]

D. Feature Importance Analysis

One of the strong features that XGBoost has built-in is Feature Importance. It gets the most important variables that are related to fraud detection which are the variables that have a significant impact on fraud detection. Early methods like Word2Vec have set the stage for semantic connections in text to be captured [13], potentially supplementing transformer-based embeddings. For more interpretable model selection, methods like SHAP values have been suggested in order to improve feature contribution explanation [5]. Alongside SHAP, model-agnostic tools like LIME offer localized interpretability of complex models [16]. CatBoost has also shown robust performance on structured datasets with categorical variables [18]. Outside of conventional ML methods, CNNs have also been studied to conduct sentence classification in NLP processes [7], while transformer architectures such as DistilBERT presently provide better understanding of context.

E. Evaluation Metrics

1) Confusion Matrix: A graphical representation of the four different situations in a confusion matrix, which are True Positives, False Positives, True Negatives, and False Negatives helps the user see what a model is about and how good or how effective it is in a given task. Table 1 illustrates a confusion matrix.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

TABLE I
CONFUSION MATRIX

2) Precision, Recall, F1-score, and AUC-ROC: Precision, Recall, F1-score and AUC-ROC are employed to analyze the performance of the model, to compare different metrics so that a balanced measurement will be attained across the board. This in turn will ensure a more comprehensive evaluation of the model's performance.

• Precision (P):Precision measures how many of the predicted positive cases were actually correct.

$$\text{Precision} = \frac{TP}{TP + FP}$$

• Recall (R):Recall measures how many of the actual positive cases were correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

• F1-Score:F1-Score is the harmonic mean of Precision and Recall. It provides a balance between the two, especially useful when data is imbalanced.

$$\text{F1} - \text{Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

• AUC-ROC (Area Under the Receiver Operating Characteristic Curve):AUC-ROC provides an aggregate measure of performance across all classification thresholds.AUC-ROC does not have a single equation but is derived from the ROC curve, which plots the following:

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

IV. RESULTS AND ANALYSIS

- **Optimized Fraud Detection Threshold:** 0.87
- **Precision:** 99.61%
- **Recall:** 100%
- **F1-score:** 99.80%
- **AUC-ROC Score:** 100%
- **Accuracy:** 99.97%

The outcomes showcase the stunning accomplishment of the suggested DistilBERT and XGBoost design. The ideal redemption rate of 100% shows that all the fake job offerings were correctly identified, thus limiting false negatives. The high accuracy rate of 99.61% is significant in ensuring that real job postings are not wrongly tagged as scam, therefore reducing false positives. Additionally, the F1-score of 99.80% is an evidence of the great balance between precision and recall. The AUC-ROC score of 100% is the main feature of the model which makes it possible to discriminate between fraudulent and genuine job postings. That's valid even if the thresholds are variable. The fraud detector's mastery of the fraud detection task is indicated by the tuned fraud detection threshold of 0.87 through a better balance of sensitivity and specificity. By these outcomes indeed, the conceived approach is practical in the real world and it could be a reliable tool for the job platforms where users would be shielded from fake activities. The low rates of errors are guarantees that the predictions are in line with what the user trusts and the platform is credible. Future research may consist in extending the dataset, implementing transfer learning for domain-specific improvement, or taking on real-time fraud detection systems.

V. VISUALIZATIONS

A. Class Distribution Before and After SMOTE

The initial dataset was plagued by a critical class imbalance, with the fraudulent job ads being significantly smaller in number than the legitimate ones. This class imbalance has the potential to create biased models biased towards the majority class. To counter this, the Synthetic Minority Oversampling Technique (SMOTE) was used. Figure 1 shows the class distribution before and after SMOTE was applied.

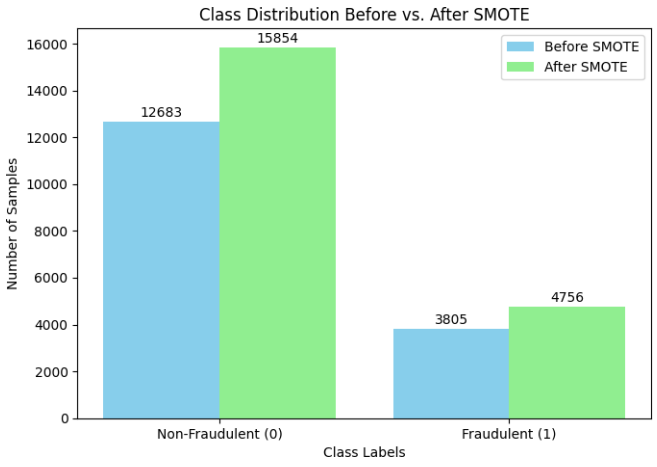

Fig. 1. Class Distribution Before and After SMOTE

B. Confusion Matrix

Confusion matrix is provided with overall model performance evaluation based on classifying test dataset. It indicates the ability of the model in distinguishing between actual and fake job postings. Figure 2 shows the confusion matrix.

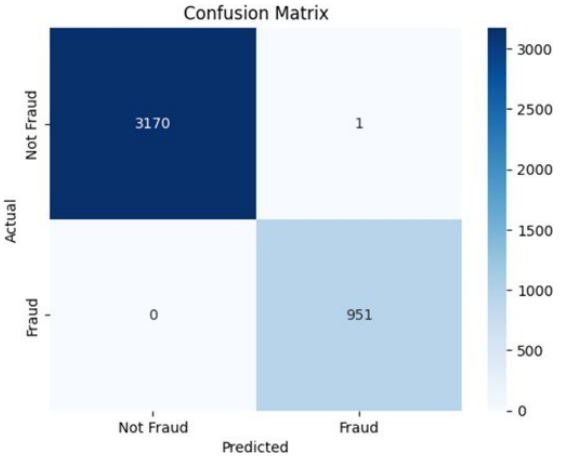

Fig. 2. Confusion Matrix

C. Feature Importance Chart

It is crucial for interpretability and trust to determine which of the features are most affecting the model's predictions. The feature importance chart illustrates the features that most affect the classification predictions. Figure 3 shows the features that has most affect on the model's predictions.

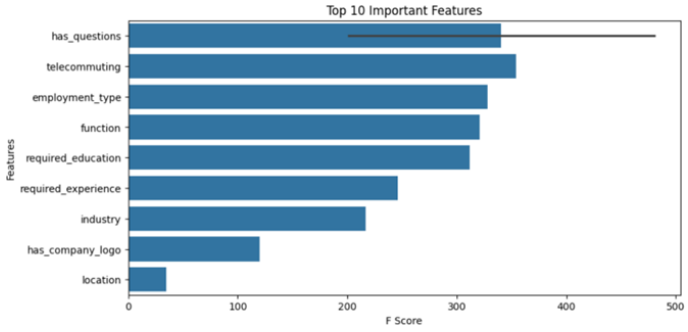

Fig. 3. Feature Importance Chart

VI. Discussion

The combined model of DistilBERT and XGBoost successfully identifies fraudulent job postings. Through feature importance analysis, missing logos of a company, vague job descriptions, and excessive benefits were revealing as primary indicators of fraud. There was a significant improvement in model performance when SMOTE was used to adjust the imbalance class. The model's high recall score shows the ability to extract fraudulent job postings with a high number of true positives and minimal false negatives. Similarly, the strong precision score ensures the number of false positives that defraud genuine job postings is minimal which is important. Through integration of this method with online job portals, the identification process can be automated to minimize manual examination and expedite fraud recognition. In addition, this will increase the security of job searchers as their chances of being taken advantage of by dishonest parties will drastically decrease. Moreover, the system's continual fusion of new information adds flexibility to the detection capabilities as well as the ever changing patterns of fraud, guaranteeing enhanced detection capabilities. Subsequent research would then be able to investigate Sentence-BERT for increased semantic comprehension on the sentence level in job postings [6]. Describing the ethical consideration of AI-based fraud detection is applicable, especially if choices would impact user confidence and opportunities [10]

VII. Study Limitations

Though the results obtained by the proposed model are positive, there are certain limitations that are relevant:

- **Dataset Scope:** The dataset covered here can be non-representative of the entire extent of global job markets or future trends of frauds. It has drawn on most of the structured and text features of available databases, which have the tendency to constrain the scope generalization to professional settings or geographically localized fraud types.
- **Static Learning:** Learned from a set data and does not have ongoing learning. As the fraudsters continue to adapt, the model can become stale without refreshing it from newer data.
- **Dated Explainability for End Users:** Even though SHAP values and feature importance are enlightening information in the direction of interpretability for engineers, interpreting them may be difficult for technical end-users or system administrators that highly rely on such systems.
- **Binary Classification Restriction:** The model addresses solely binary classification (legitimate or fraudulent), which may be overly theoretical for the case where postings are suspicious but not necessarily fraudulent, and maybe fails to encompass edge cases.
- **Dependent upon Text Quality:** Because the model relies greatly on NLP to search for text features, grammar errors, local spelling, or deliberately obfuscatory use of employment postings can be effective.

VII. CONCLUSION AND FUTURE SCOPE

This paper develops a strong AI and machine learning-powered deception job posting detection model using Natural Language Processing (NLP). The methodology is accurate and portable and presents a usable way of detecting web-based job scams. Areas of future research would involve scaling the dataset, model explainability, utilizing the model as a CTFD fraud detection API, and adding trust and transparency. Additionally, integrating the system into a series of job portals and adaptive ongoing learning would continue to improve the model's ability to detect novel schemes of fraud. In addition, the application of interpretability methods like SHAP or LIME can improve stakeholder trust through trans- parency of explanation to decisions, which is paramount for ethical use of AI. Real-time feature detection ability and compatibility with security infrastructure can improve the model's utility in dynamic settings as well. Fraudsters adapt strategies to counter them, and therefore, the inclusion of a feedback-induced retraining function will be vital in keeping the system up to date with new emerging fraudulent schemes. The contributions of the work are not only adding to the secure online hiring body of

knowledge but also opening the door towards the creation of intelligent trust-based screening systems in other domains vulnerable to digital fraud. To further enhance the strength and reusability of the pro- posed framework, some directions for future research include a Real-Time Detection Framework by designing a real-time fraud detection framework that can be directly implemented in recruitment portals and dynamically adapt its predictions based on the postings executed. Also, Multinational and Multi- lingual Dataset Augmentation can be implemented where more advanced language, language contained dataset will further enhance the detection of fraud within other geographies and labor markets. Image and Metadata Feature Utilization by subsequent models can be implemented which use firm logos, recruiter profile images, and domain registration as features and have a stronger detection of fraud relative to text and simple attributes. Transfer Learning and Fine-Tuning can be implemented by the expert models (e.g., legal and medical ex- perts) which may be used for transfer learning to enhance fraud detection for expert career types. Enhanced explainability to end-users can be done by creating simple-to-use dashboards that incorporate simple-to-understand english explanations for fraud risk which will likely make recruiters and applicants trust and employ them. Also DistilBERT and XGBoost can be ensembled with other top-performing classifiers like CatBoost, LSTM, or CNN to further improve detection performance for different types of postings.

REFERENCES

[1] H. Tabassum, G. Ghosh, A. Atika and A. Chakrabarty, ”Detecting Online Recruitment Fraud Using Machine Learning,” 2021 9th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 2021, pp. 472-477, doi: 10.1109/ICoICT52021.2021.9527477

[2] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” Neural Netw., vol. 106, pp. 249–259, 2018. DOI: 10.1016/j.neunet.2018.07.011.

[3] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter,” arXiv preprint arXiv:1910.01108, 2019. [Online]. Available: https://arxiv.org/abs/1910.01108.

[4] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “LightGBM: A highly efficient gradient boosting decision tree,” in Adv. Neural Inf. Process. Syst., vol. 30, 2017. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

[5] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Adv. Neural Inf. Process. Syst., vol. 30, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

[6] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in Proc. EMNLP-IJCNLP, 2019. [Online]. Available: https://arxiv.org/abs/1908.10084

[7] Y. Zhang and B. C. Wallace, “A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification,” arXiv preprint arXiv:1510.03820, 2015. [Online]. Available: https://arxiv.org/abs/1510.03820.

[8] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785–794. Available: arXiv:1603.02754.

[9] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, “Data mining for credit card fraud: A comparative study,” Decis. Support Syst., vol. 50, no. 3, pp. 602–613, 2011. DOI: 10.1016/j.dss.2010.08.008.

[10] R. Binns, M. Veale, M. Van Kleek, and N. Shadbolt, “‘It’s reducing a human being to a percentage’: Perceptions of justice in algorithmic decisions,” in Proc. 2018 CHI Conf. Human Factors in Comput. Syst., 2018, p. 377. Available: arXiv:1801.10408.

[11] B. Krawczyk, “Learning from imbalanced data: Open challenges and future directions,” Prog. Artif. Intell., vol. 5, no. 4, pp. 221–232, 2016. DOI: 10.1007/s13748-016-0094-0.

[12] N. Japkowicz and S. Stephen, “The class imbalance problem: A systematic study,” Intelligent Data Analysis, vol. 6, no. 5, pp. 429–449, 2002. DOI: 10.3233/IDA-2002-6504.

[13] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013. [Online]. Available: https://arxiv.org/abs/1301.3781.

[14] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, 1997. DOI: 10.1006/jcss.1997.1504.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019. [Online]. Available: https://arxiv.org/abs/1810.04805.

[16] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You?: Explaining the Predictions of Any Classifier,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD), New York, NY, USA, 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[17] H. He and E. A. Garcia, “Learning from imbalanced data,” IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, 2009. DOI: 10.1109/TKDE.2008.239.

[18] L. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, and A. Gulin, “CatBoost: Unbiased boosting with categorical features,” in Adv. Neural Inf. Process. Syst., vol. 31, 2018. [Online]. Available: https://arxiv.org/abs/1706.09516.