

ABSTRACT

Most companies nowadays are using digital platforms for the recruitment of new employees to make the hiring process easier, faster, and more accessible. The rapid increase in the use of online platforms for job posting has resulted in a significant rise in fraudulent advertising and malicious job scams. The scammers are making money through fraudulent job postings by luring unsuspecting job seekers with attractive, yet fake, employment offers. As a result, online recruitment fraud has emerged as an important issue in the domain of cybercrime, causing financial and emotional harm to job seekers. Therefore, it is necessary to detect fake job postings to protect users and get rid of online job scams. In recent studies, both traditional machine learning and advanced deep learning algorithms have been implemented to detect fake job postings. This research aims to use two state-of-the-art transformer-based deep learning models, i.e., Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT-Pretraining Approach (RoBERTa), to detect fake job postings more precisely and efficiently. In this research, a novel dataset of fake job postings is proposed, formed by the combination of job postings from three different sources to ensure diversity and relevance. Existing benchmark datasets are outdated and limited due to knowledge of specific job postings, which restricts the existing models' capability in detecting newly evolving fraudulent job advertisements. Hence, we extend it with the latest job postings collected from trusted sources. Exploratory Data Analysis (EDA) highlights a noticeable class imbalance problem in detecting fake jobs, which tends to make the model biased and act aggressively toward the minority class. In response to overcome this issue, the work at hand implements ten top-performing variants of the Synthetic Minority Oversampling Technique (SMOTE), a popular oversampling method for handling imbalanced datasets. The models' performances balanced by each SMOTE variant are thoroughly analyzed and compared through multiple evaluation metrics. All implemented approaches perform competitively in identifying fraudulent jobs. However, the combination of BERT and SMOBD SMOTE achieved the highest balanced accuracy and recall of about 90%, showcasing its effectiveness in fake job detection tasks.