

ABSTRAK

Maraknya Online Recruitment Fraud (ORF) menjadi ancaman bagi pencari kerja, terutama pada lowongan kerja yang tersebar dalam bentuk teks maupun poster digital. Penelitian ini membangun website Scanix untuk mendeteksi lowongan kerja palsu dengan menggabungkan analisis teks, analisis visual, dan validasi kontekstual. Metode yang digunakan meliputi OCR untuk mengekstraksi teks dari poster, TF-IDF dan SVM untuk klasifikasi teks, SVM visual untuk fitur citra, serta CNN VGG16 untuk analisis gambar. Hasil prediksi dari model digabungkan menggunakan decision-level fusion dan diperkuat dengan validasi email serta URL. Sistem dikembangkan menggunakan arsitektur client-server dengan Vue.js dan Flask. Hasil evaluasi menunjukkan tingkat akurasi untuk masing-masing model yaitu: SVM Teks (92,75%), SVM Visual (82,28%), CNN VGG16 (94,94%), dan model Ensemble memperoleh akurasi 90,88% dengan AUC 0,9632. Perbedaan skor ini mengindikasikan bahwa fitur spasial (CNN) dan linguistik (SVM Teks) sangat dominan dalam mendeteksi penipuan, di mana pendekatan Ensemble hadir untuk menyatukan kekuatan tersebut sekaligus menutupi kelemahan fitur visual manual. Pengujian sistem menunjukkan 176 assertion Backend API berhasil, 42 skenario End-to-End berhasil, serta skor Lighthouse Mobile dan Desktop memperoleh Performance 98, Accessibility 95, Best Practices 100, dan SEO 100.

Kata Kunci: *Convolutional Neural Network, Ensemble Learning, Online Recruitment Fraud, Support Vector Machine, Validasi Kontekstual.*

ABSTRACT

The increasing prevalence of Online Recruitment Fraud (ORF) poses risks to job seekers, especially when fake vacancies are distributed as digital text or job posters. This study develops Scanix, a web-based system for detecting fake job vacancies by combining text analysis, visual analysis, and contextual validation. The proposed method applies OCR to extract text from posters, TF-IDF and SVM for text classification, visual SVM for image features, and CNN VGG16 for image-based analysis. The prediction results are combined using decision-level fusion and supported by email and URL validation. The system is implemented using a client-server architecture with Vue.js and Flask. The evaluation results showed the accuracy levels for each model as follows: Text SVM (92.75%), Visual SVM (82.28%), CNN VGG16 (94.94%), and the Ensemble model achieved 90.88% accuracy with an AUC of 0.9632. This performance variance indicates that spatial features (CNN) and linguistic features (Text SVM) are highly dominant in detecting fraud, with the Ensemble approach uniting these strengths to overcome the limitations of manual visual features. System testing showed that 176 Backend API assertions passed, 42 End-to-End scenarios passed, and Lighthouse scores on both Mobile and Desktop reached Performance 98, Accessibility 95, Best Practices 100, and SEO 100.

Keywords: Contextual Validation, Convolutional Neural Network, Ensemble Learning, Online Recruitment Fraud, Support Vector Machine.