

ABSTRAK

Proses penyaringan *Curriculum Vitae* (CV) dalam rekrutmen bidang Teknologi Informasi memerlukan penilaian hubungan semantik antara profil pelamar dan deskripsi pekerjaan. Perbedaan struktur, bahasa, dan istilah teknis membuat pencocokan manual memerlukan waktu serta sulit dilakukan secara konsisten. Penelitian ini mengimplementasikan Multilingual BERT (mBERT) dengan pendekatan *cross-encoder* untuk klasifikasi biner dan *ranking* kecocokan CV terhadap *job description* (JD). Tahapan penelitian mengikuti CRISP-DM yang mencakup pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi, dan pengujian inferensi melalui prototipe Streamlit. *Dataset* terdiri atas 500 CV dan 2.264 JD dalam 10 *role family* Teknologi Informasi. Setiap *role* memiliki 50 CV dengan komposisi 25 CV berbahasa Indonesia dan 25 CV berbahasa Inggris. *Skill-aware pairing* menghasilkan 12.500 pasangan, terdiri atas 10 pasangan positif dan 15 pasangan negatif untuk setiap CV. Data dibagi berdasarkan *resume_id* menjadi 8.750 pasangan latih, 1.875 pasangan validasi, dan 1.875 pasangan uji untuk mencegah *data leakage*. Model *bert-base-multilingual-cased* dilatih menggunakan *partial fine-tuning* dengan membekukan *embedding layer* dan *encoder layer* 0-3. *Checkpoint* terbaik diperoleh pada *epoch* ke-8 dengan *validation macro-F1* 0,8182. Pada *test set* dan *threshold* 0,50, model menghasilkan *accuracy* 0,8448, *F1-score* 0,8124, *macro-F1* 0,8400, dan ROC-AUC 0,9216. Evaluasi *ranking* menghasilkan MRR 0,9411, *Average Precision* 0,8795, Hit@1 0,9067, dan Hit@10 1,0000.

Kata Kunci: *Curriculum Vitae*, *Job Description*, Multilingual BERT, *Cross-Encoder*, CRISP-DM.

ABSTRACT

Curriculum Vitae (CV) screening in Information Technology recruitment requires an assessment of the semantic relationship between applicant profiles and job descriptions. Differences in structure, language, and technical terminology make manual matching time-consuming and difficult to perform consistently. This study implements Multilingual BERT (mBERT) with a cross-encoder approach for binary classification and ranking of CV-job description (JD) matches. The research follows CRISP-DM, covering business understanding, data understanding, data preparation, modeling, evaluation, and inference testing through a Streamlit prototype. The dataset contains 500 CVs and 2,264 JDs across 10 Information Technology role families. Each role contains 50 CVs, comprising 25 Indonesian CVs and 25 English CVs. Skill-aware pairing produced 12,500 pairs, with 10 positive and 15 negative pairs for each CV. The data were split by resume_id into 8,750 training pairs, 1,875 validation pairs, and 1,875 test pairs to prevent data leakage. The bert-base-multilingual-cased model was trained using partial fine-tuning by freezing the embedding layer and encoder layers 0-3. The best checkpoint was obtained at epoch 8 with a validation macro-F1 of 0.8182. On the test set at a 0.50 threshold, the model achieved an accuracy of 0.8448, F1-score of 0.8124, macro-F1 of 0.8400, and ROC-AUC of 0.9216. Ranking evaluation produced an MRR of 0.9411, Average Precision of 0.8795, Hit@1 of 0.9067, and Hit@10 of 1.0000.

Keywords: Curriculum Vitae, Job Description, Multilingual BERT, Cross-Encoder, CRISP-DM.