

INTISARI

OPTIMASI MODEL REKOMENDASI PEKERJAAN MELALUI PENINGKATAN ROBUSTNESS VEKTOR EMBEDDING LOWONGAN KERJA DAN RESUME BERBASIS TSDAE

Oleh

Muhammad Adin Palimbani
23/530097/PPA/06713

Model rekomendasi pekerjaan berbasis kecerdasan buatan masih menghadapi tantangan besar dalam menangani struktur informasi lowongan kerja yang tidak berlabel, tidak konsisten dan *noisy*. Judul dan kategori pekerjaan yang diunggah oleh perusahaan sering kali bervariasi, menyebabkan *overlapping content* dan kesulitan dalam klasifikasi otomatis. Penelitian ini bertujuan mengoptimalkan model rekomendasi pekerjaan adaptif melalui peningkatan *robustness* vektor embedding lowongan kerja dan resume terhadap data *noisy*, *inconsistent*, *unstructured* dan *unlabelled* menggunakan *Transformer based Sequential Denoising Auto-Encoder* (TSDAE). Proses pembentukan embedding akan dilakukan secara terintegrasi melalui 3 tahapan *Domain Adaptive Pre-Training* yakni menggunakan *Pre-Trained Language* BERT (baseline), *training* TSDAE pada domain target dan *fine-tuning* hasil TSDAE *pre-trained*. Dataset yang digunakan adalah 4000 informasi lowongan kerja dari Kaggle, O*NET *Standard Occupation* dan lima Resume sebagai skenario pengujian. Hasil *embedding* kemudian dikelompokkan dengan *K-Means Clustering* dengan $k=12$ dan diukur kemiripannya menggunakan *Cosine Similarity*. Setiap pekerjaan yang direkomendasikan sistem akan dianotasi oleh lima annotator. Model dievaluasi pada Top-20 Rekomendasi dengan metrik MAP@20, NDCG@20, P@20 dan MRR@20. Model yang diusulkan menunjukkan bahwa TSDAE *fine-tuned* secara signifikan mengungguli model *baseline* BERT tanpa TSDAE, dengan mencapai skor rata-rata MAP@20 86%, NDCG@20 93%, P@20 80% dan MRR@20 90%. Hasil ini membuktikan pendekatan TSDAE yang dioptimalkan mampu menghasilkan rekomendasi pekerjaan yang adaptif dan relevan terhadap karakteristik data *noisy*, *inconsistent* dan *unlabelled*.

Kata Kunci: Rekomendasi Pekerjaan, Domain Adaptation, TSDAE, *K-Means Clustering*, *Cosine Similarity*

ABSTRACT

OPTIMIZATION OF JOB RECOMMENDATION MODELS THROUGH ENHANCING THE ROBUSTNESS OF JOB VACANCY AND RESUME EMBEDDING VECTORS USING TSDAE

Oleh

Muhammad Adin Palimbani

23/530097/PPA/06713

Artificial intelligence-based job recommendation models still face major challenges in handling unlabeled, inconsistent, and noisy job vacancy information structures. Job titles and categories uploaded by companies often vary, causing overlapping content and difficulties in automatic classification. This research aims to optimize adaptive job recommendation models by improving the robustness of job vacancy and resume embedding vectors against noisy, inconsistent, unstructured, and unlabeled data using Transformer-based Sequential Denoising Auto-Encoder (TSDAE). The embedding formation process will be carried out in an integrated manner through three stages of Domain Adaptive Pre-Training, namely using Pre-Trained Language BERT (baseline), training TSDAE on the target domain, and fine-tuning the results of TSDAE pre-trained. The dataset used is 4,000 job vacancy information from Kaggle, O*NET Standard Occupation, and five resumes as a test scenario. The embedding results are then grouped using K-Means Clustering with $k=12$ and their similarity is measured using Cosine Similarity. Each job recommended by the system will be annotated by five annotators. The model is evaluated on the Top-20 Recommendations using the metrics $MAP@20$, $NDCG@20$, $P@20$, and $MRR@20$. The proposed model demonstrates that TSDAE *fine-tuned* significantly outperforms the *baseline* BERT model without TSDAE, achieving an average score of $MAP@20$ 86%, $NDCG@20$ 93%, $P@20$ 80%, and $MRR@20$ 90%. These results demonstrate that the optimized TSDAE approach is capable of generating adaptive and relevant job recommendations for noisy, inconsistent, and unlabeled data.

Keywords: Job Recommendation Model, Domain Adaptation, TSDAE, K-Means Clustering, Cosine Similarity