INTISARI

ANALISIS KOMPARATIF MODEL *TRANSFORMER* (INDOBERT, XLNET, DAN DISTILBERT) UNTUK PEMETAAN OTOMATIS *SCOPE OF SUPPLY* PADA PENGADAAN JASA PT PUPUK INDONESIA (PERSERO)

Oleh

Cornelia Happy Rahmawati
22/496781/PA/21362

Penelitian ini bertujuan mengatasi permasalahan misklasifikasi dalam proses pengadaan jasa di PT Pupuk Indonesia (Persero) akibat tingginya granularitas *Scope of Supply* (SoS) dan perbedaan interpretasi antar pemangku kepentingan. Pendekatan yang digunakan adalah pemetaan otomatis Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) ke dalam SoS melalui analisis komparatif model *Transformer*, yaitu IndoBERT, XLNet, dan DistilBERT. Data penelitian terdiri atas 398 Item *Service Code* (ISC), 1.790 kode KBLI 2020, serta data *ground truth* dari riwayat Purchase Order periode 2021–2025. Metode yang diterapkan meliputi text *Preprocessing*, ekstraksi *embedding* dengan parameter seragam, serta perhitungan *cosine similarity* untuk menghasilkan peringkat kandidat KBLI. Evaluasi dilakukan menggunakan metrik Top-5, yaitu Hit@5, *Precision*@5, *Recall*@5, F1@5, dan mean Reciprocal Rank (MRR). Hasil penelitian menunjukkan bahwa XLNet memiliki performa terbaik, diikuti oleh IndoBERT dan DistilBERT, sementara DistilBERT unggul dalam efisiensi komputasi. Berdasarkan hasil tersebut, XLNet direkomendasikan sebagai model paling optimal karena memberikan keseimbangan terbaik antara akurasi dan efisiensi dalam mendukung proses pengadaan jasa.

Kata kunci: *Information retrieval*, model *Transformer*, IndoBERT, XLNet, DistilBERT, *semantic similarity*, pengadaan jasa.

ABSTRACT

A COMPARATIVE ANALYSIS OF TRANSFORMER MODELS (INDOBERT, XLNET, AND DISTILBERT) FOR AUTOMATIC SCOPE OF SUPPLY MAPPING IN SERVICE PROCUREMENT AT PT PUPUK INDONESIA (PERSERO)

By

Cornelia Happy Rahmawati
22/496781/PA/21362

This study aims to address misclassification issues in the service procurement process at PT Pupuk Indonesia (Persero), which arise due to the high granularity of the Scope of Supply (SoS) and differing interpretations among stakeholders. The proposed approach involves automatically Mapping the Indonesian Standard Industrial Classification (KBLI) to the SoS using a comparative analysis of Transformer-based models, namely IndoBERT, XLNet, and DistilBERT. The dataset consists of 398 Item Service Codes (ISC), 1,790 KBLI 2020 codes, and ground truth data derived from Purchase Order records for the period 2021–2025. The methodology includes text Preprocessing, embedding extraction with uniform parameter settings, and cosine similarity computation to generate ranked KBLI candidates. Model performance is evaluated using Top-5 ranking-based metrics, including Hit@5, Precision@5, Recall@5, F1@5, and Mean Reciprocal Rank (MRR). The results indicate that XLNet achieves the best performance, followed by IndoBERT and DistilBERT, while DistilBERT demonstrates superior computational efficiency. Based on these findings, XLNet is recommended as the most optimal model, as it provides the best balance between Accuracy and efficiency in supporting the service procurement process.

Keywords: *Information retrieval, Transformer model, IndoBERT, XLNet, DistilBERT, semantic similarity, service Procurement.*