

ABSTRAK

Artayasa, Ketut (2025), *Prediksi Kode Kelompok Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) dengan Deep Learning*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Kata Kunci : Klasifikasi, KBLI, IndoBERT, IndoBERTkbli

Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) adalah pengklasifikasian aktivitas ekonomi di Indonesia yang menghasilkan produk atau output baik berupa barang maupun jasa yang disusun berdasarkan lapangan usaha. KBLI mengelompokkan kegiatan ekonomi ke dalam struktur yang konsisten dan saling berhubungan berdasarkan konsep, definisi, prinsip, dan tata cara pengklasifikasian yang telah disepakati secara internasional. Penentuan kode KBLI sering kali menghadapi tantangan, seperti variasi interpretasi deskripsi usaha, ketidaksesuaian bahasa, dan proses manual yang tidak efisien serta rentan terhadap kesalahan. Sistem pencarian KBLI saat ini masih menggunakan *exact matching key*, sehingga sangat tergantung dengan ketepatan kata kunci, ditambah lagi bahwa kata kunci yang sama ada di beberapa KBLI yang berbeda. Penentuan kode KBLI terutama kode 5-digit menjadi tantangan tersendiri, karena kode harus dicari yang sesuai dengan deskripsi usaha atau aktivitas ekonomi, selain membutuhkan waktu, juga dituntut kehati-hatian agar tidak salah dalam menentukan kode KBLI yang sesuai. Menyadari pentingnya kesesuaian isian kode KBLI, penelitian ini mengajukan prediksi kode kelompok KBLI 5-digit dengan model *deep learning*. Melalui model yang dikembangkan pada penelitian ini diharapkan dapat memprediksi berdasarkan konteks input deskripsi aktivitas ekonomi dan produknya berada pada kelompok klasifikasi KBLI yang sesuai, penelitian dilakukan dengan melakukan *hyperparameter fine-tuning* pada model *IndoBERT* yang sudah memahami konteks Bahasa Indonesia, pada penelitian ini juga dilakukan augmentasi sinonim pada dataset dengan sampel dibawah 20, yang bertujuan untuk mengatasi masalah jumlah sampel kelas yang tidak seimbang, sekaligus diharapkan dapat meningkatkan akurasi.

ABSTRACT

Artayasa, Ketut (2025), *Prediction of Indonesian Standard Industrial Classification (KBLI) Group Codes Using Deep Learning*. Thesis, Computer Science, Postgraduate Program, Ganesha University of Education.

Keywords : Classification, KBLI, IndoBERT, IndoBERTkbli

The Indonesian Standard Industrial Classification (KBLI) is a classification of economic activities in Indonesia that produce products or outputs, either in the form of goods or services, organized based on business fields. KBLI groups economic activities into a consistent and interconnected structure based on internationally agreed concepts, definitions, principles, and classification procedures. Determining KBLI codes often faces challenges, such as variations in interpreting business descriptions, language mismatches, and inefficient manual processes prone to errors. The current KBLI search system relies on exact matching keys, making it highly dependent on keyword accuracy, compounded by the issue of identical keywords appearing in multiple KBLI categories. Determining the 5-digit KBLI code is particularly challenging, as it requires identifying the code that matches the business description or economic activity, which is time-consuming and demands careful consideration to avoid errors. Recognizing the importance of accurate KBLI code assignment, this study proposes predicting 5-digit KBLI group codes using a deep learning model. The developed model is expected to predict the appropriate KBLI classification group based on the context of the input description of economic activities and their products. The study employs hyperparameter fine-tuning on the IndoBERT model, which understands the context of the Indonesian language. Additionally, synonym augmentation is applied to datasets with fewer than 20 samples to address the issue of imbalanced class sample sizes, aiming to improve accuracy.