

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Penelitian**

Klasifikasi Baku Lapangan Indonesia (KBLI) merupakan klasifikasi aktivitas/kegiatan ekonomi Indonesia yang menghasilkan produk/output, baik berupa barang maupun jasa, berdasarkan lapangan usaha yang digunakan sebagai acuan standar dan alat koordinasi, integrasi, serta sinkronisasi penyelenggaraan statistik. KBLI merupakan salah satu pertanyaan sangat penting pada Survei dan Sensus Ekonomi yang dilaksanakan oleh Badan Pusat Statistik (BPS), yang merupakan sumber informasi terkait karakteristik suatu usaha menurut lapangan usaha dan produk yang dihasilkan (Badan Pusat Statistik, 2020).

Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) pada awalnya dirancang untuk memenuhi keperluan analisis ekonomi, pengambilan keputusan, dan pembuatan kebijakan. Saat ini KBLI juga digunakan untuk penentuan bidang usaha yang tercantum pada *Online Single Submission* (OSS) yaitu sistem yang digunakan untuk mempermudah pengajuan perizinan usaha secara daring.

Ketepatan dalam menentukan kode KBLI sangat menentukan kualitas data dan analisis ekonomi yang dihasilkan. Banyaknya klasifikasi dalam KBLI menyebabkan petugas memerlukan waktu relatif lama untuk menentukan kode KBLI suatu aktivitas ekonomi. Selain karena banyaknya kelompok KBLI, beberapa aktivitas ekonomi yang hampir mirip berada pada kelompok KBLI yang berbeda, kesalahan dalam memberikan kode KBLI akan berdampak pada kesalahan kesimpulan terhadap data yang dihasilkan.

Saat ini pencarian kode KBLI walaupun sudah memanfaatkan teknologi informasi, namun oleh karena hanya menggunakan pencarian biasa dengan metode *exact match key*, pencarian hanya akan efektif jika kode KBLI dicari dengan kata kunci yang sesuai. Hal ini menyebabkan pencarian kode KBLI memerlukan waktu relatif lama sampai dengan menemukan kode KBLI yang benar-benar sesuai dengan kegiatan ekonomi dari suatu usaha.

Selain penentuan kode, saat ini juga belum ada sistem pemeriksaan KBLI yang bisa dilakukan secara otomatis, pemeriksaan dilakukan dengan cara manual, yaitu melakukan pengecekan antara isian kode KBLI, deskripsi aktivitas usaha, produk usaha dengan deskripsi kegiatan ekonomi yang ditetapkan pada buku KBLI. Kegiatan ini juga membutuhkan waktu yang relatif lama, apalagi untuk kegiatan dengan skala besar, misalnya Sensus Ekonomi.

Kemajuan teknologi *machine learning*, dalam hal ini terutama *deep learning* memiliki potensi untuk memberikan solusi yang lebih efisien. Kemampuan *deep learning* dalam memproses data berupa teks terus meningkat, mampu memahami pola kompleks, dan terbukti efektif dalam tugas klasifikasi pada berbagai domain, termasuk pengolahan bahasa alami.

Sebagai percobaan awal dengan menggunakan 21.926 baris deskripsi aktivitas ekonomi dan produknya, dilakukan prediksi kode KBLI dengan menggunakan model GPT-4.1 diperoleh akurasi 12,59 persen, akurasi yang cukup rendah untuk bisa dijadikan model prediksi kode kelompok KBLI.

Berdasarkan latar belakang di atas, penelitian ini mengajukan prediksi kode kelompok KBLI dengan *deep learning* yang dapat dimanfaatkan untuk

mengotomatisasi proses penentuan kode KBLI, meningkatkan akurasi, dan mempercepat pengelompokan data usaha, serta melakukan evaluasi ketepatan kode KBLI pada data hasil penyelenggaraan statistik. Adapun penelitian yang diajukan berjudul “**Prediksi Kode Kelompok Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) dengan *Deep Learning***”.

## 1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang di atas, penulis mengidentifikasi beberapa masalah dalam penelitian ini sebagai berikut.

1. Kode KBLI sangat penting dalam analisis ekonomi, diperlukan mekanisme penentuan kode yang lebih cepat dan tepat berbasis *machine learning*.
2. Penentuan dan pemeriksaan kode kelompok KBLI masih dilakukan secara manual membutuhkan waktu lama dan masih rentan terjadi kesalahan, sangat tergantung pemahaman petugas terhadap konsep tertentu. Sistem pencarian kode KBLI saat ini menggunakan *exact matching key*, sehingga sangat ditentukan dengan ketepatan kata kunci yang digunakan.

## 1.3 Pembatasan Masalah

Untuk mempermudah fokus penelitian, penulis menetapkan batasan masalah yang akan dibahas sebagai berikut :

1. Untuk pembuatan model KBLI, penelitian dilakukan dengan melakukan *hyperparameter fine-tuning* dengan *base model IndoBERT (Indobert-base-p2)*.  
*Hyperparameter fine-tuning* dilakukan pada parameter:

- *Learning rate*:  $1 \times 10^{-5}$ ,  $2 \times 10^{-5}$ , dan  $3 \times 10^{-5}$ , menurut penelitian (Sun dkk., 2020) *learning rate* kecil lebih efektif untuk tugas klasifikasi yang kompleks.
  - *Batch size*: 16 dan 32
  - *Number of epochs*: 10
2. Dataset bersumber dari data Sensus Ekonomi 2016 dengan lokus Provinsi Bali dengan basis KBLI-2015.
  3. Pada penelitian ini juga dilakukan augmentasi data pada kelas yang jumlah sampelnya dibawah 20 sampel, dengan cara melakukan variasi berbasis sinonim, dengan mengganti beberapa kata yang dapat menambah jumlah dataset dengan makna yang hampir sama, misalnya kata “menjual” diganti dengan “berdagang”, kata “membuat” diganti dengan kata “memproduksi”.
  4. KBLI yang digunakan pada dataset adalah KBLI 2015 yang sudah dipadankan dengan KBLI2020 pada lokus Provinsi Bali.
  5. Kelas yang dilatih dalam proses *fine-tuning* adalah kelas dengan jumlah data minimal 20 data setelah melalui tahapan augmentasi.
  6. Kode KBLI yang diprediksi adalah kode 5-digit kelompok KBLI.

#### 1.4 Rumusan Masalah

Berdasarkan masalah yang telah diidentifikasi sebelumnya, rumusan masalah dalam penelitian ini dapat dirinci sebagai berikut :

1. Kombinasi hiperparameter yang mana yang paling optimal dalam *fine-tuning* model klasifikasi multi-kelas Klasifikasi Baku Lapangan Usaha Indonesia?

2. Apakah model layak menjadi model prediksi kode kelompok Klasifikasi Baku Lapangan Usaha Indonesia?

### **1.5 Tujuan Penelitian**

Tujuan dari penelitian ini meliputi:

- 1 Mendapatkan kombinasi hiperparameter yang paling optimal.
- 2 Membuat model untuk memprediksi kode kelompok KBLI dengan melakukan *hyperparameter fine-tuning* pada *base model* IndoBERT.

### **1.6 Manfaat Penelitian**

Diharapkan bahwa temuan dari penelitian ini akan memberikan manfaat secara akademis dan praktis, yaitu :

#### **1.6.1 Manfaat Akademik**

Manfaat akademik yaitu hasil penelitian ini diharapkan dapat memberikan manfaat bagi pengembangan ilmu pengetahuan khususnya dalam hal pemilihan hiperparameter yang optimal pada klasifikasi multi-kelas kode kelompok KBLI, augmentasi sinonim pada aktivitas ekonomi dan produknya dan menjadi model *baseline* untuk penelitian prediksi kode kelompok KBLI.

#### **1.6.2 Manfaat praktis**

Model hasil penelitian ini dapat dimanfaatkan untuk diimplementasikan pada sistem prediksi, validasi, dan pemeriksaan kode KBLI secara otomatis.