

SISTEM AUTOMASI KLASIFIKASI INDUSTRI PADA KLASIFIKASI BAKU LAPANGAN USAHA INDONESIA (KBLI)

D.N.A.A.A. Mahandiri¹, A. Muliantara²

ABSTRAK

Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) adalah sebuah sistem pengkodean yang digunakan untuk mengklasifikasikan berbagai jenis kegiatan usaha di Indonesia yang berperan penting dalam pendataan dan pengawasan jenis usaha. Namun proses pengklasifian ini cukup sulit dilakukan dan masih memakan waktu lama karena kurangnya pengetahuan ataupun kelasan manual dari sumber daya manusia sehingga diperlukan sebuah sistem otomatis yang mampu mengklasifikasikan KBLI dengan mudah dan cepat dengan memanfaatkan algoritma deep learning yaitu IndoBERT yang terintegrasi dalam sebuah website. Advanced Indonesian Standard Classification Industrial Assistance (ASICA) sebagai sistem yang mampu menyelesaikan permasalahan klasifikasi 3 kode kategori dari KBLI 1-digit berbasis website.

Kata kunci : Klasifikasi teks, KBLI, IndoBert, Deep Learning

ABSTRACT

The Indonesian Standard Industrial Classification (KBLI) is a coding system used to classify various types of business activities in Indonesia. It plays a crucial role in data collection and monitoring of business types. However, the classification process is quite challenging and time-consuming due to the lack of knowledge or manual classification by human resources. Therefore, an automated system is needed to classify KBLI easily and quickly by utilizing deep learning algorithms, specifically, IndoBERT, integrated into a website. The Advanced Indonesian Standard Classification Industrial Assistance (ASICA) is a system capable of addressing the classification issues of the 3-digit category codes in KBLI, based on a website platform.

Keywords: Text Classification, KBLI, IndoBert, Deep Learning

1. PENDAHULUAN

Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) adalah sebuah sistem pengkodean yang digunakan untuk mengklasifikasikan berbagai jenis kegiatan usaha di Indonesia. KBLI memainkan

¹ Program Studi Informatika, Fakultas MIPA, Universitas Udayana, adiagung707@gmail.com

² Program Studi Informatika, Fakultas MIPA, Universitas Udayana, muliantara@unud.ac.id

Submitted: 7 Oktober 2023

Revised: 26 Oktober 2023

Accepted: 26 Oktober 2023

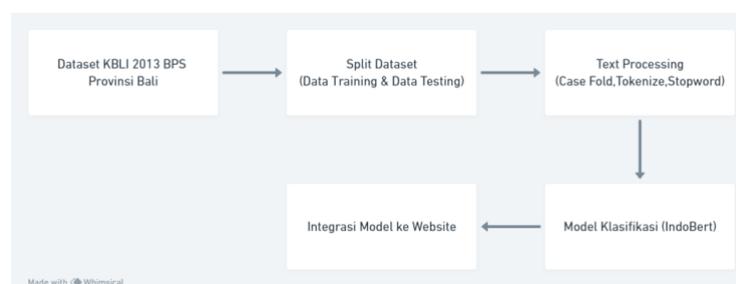
peran penting dalam pengumpulan data ekonomi, analisis statistik, serta pemantauan dan pengaturan sektor usaha di Indonesia. Setiap jenis usaha dalam KBLI diberikan kode kategori yang unik, yang membantu dalam identifikasi, klasifikasi, dan pemahaman aktivitas usaha yang dilakukan (BPS, 2023)

Namun, proses klasifikasi kode kategori dalam KBLI secara manual dapat menjadi tugas yang kompleks dan memakan waktu, serta kerap kali menimbulkan kesalahan dalam pendataan ataupun kesulitan dalam mendata karena kurangnya pengetahuan terutama terkait Klasifikasi Baku Lapangan Usaha Indonesia (KBLI). Untuk mempermudah dalam mengelola data dan meningkatkan efisiensi waktu dalam mengkategorikan diperlukan sebuah sistem yang dapat memilah pengaduan secara otomatis sesuai dengan deskripsi dan cakupan usaha pada setiap kategori. Konsep text mining dapat digunakan untuk menyelesaikan permasalahan seperti ini. Text mining menerapkan konsep dari teknik data mining untuk mencari pola menarik dalam teks yang

dapat digunakan untuk klasifikasi teks (Zong et al., 2021). Text mining mengolah data tidak terstruktur. Dalam prosesnya, diperlukan tahap awal (text preprocessing) untuk mempersiapkan teks menjadi lebih terstruktur sehingga siap diolah dengan algoritma tertentu. Proses klasifikasi dilakukan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas untuk kepentingan tertentu (Sidik et al., 2020). Metode klasifikasi yang digunakan tentunya harus mampu mengklasifikasikan teks dalam bahasa Indonesia. Beberapa penelitian dalam kurun 3 tahun diantaranya membahas tentang kemampuan metode klasifikasi dengan algoritma IndoBERT (Koto et al., 2020). IndoBERT itu sendiri terdiri dari BERT yaitu Bidirectional Encoder Representations from Transformers yang biasa digunakan untuk masalah deep learning (Juarto & Yulianto, 2023). Pengembangan sistem klasifikasi menggunakan algoritma IndoBERT telah dilakukan dalam untuk mengklasifikasikan opini publik terhadap vaksin Covid-19 di Indonesia (Saadah et al., 2022). Pelaksanaan ini bertujuan untuk mengklasifikasikan kode kategori usaha Perusahaan berdasarkan KBLI dengan memanfaatkan algoritma IndoBert yang disajikan dalam bentuk user interface website sederhana.

2. METODE PELAKSANAAN

Untuk mencapai tujuan pengembangan menggunakan sistem klasifikasi KBLI otomatis untuk Badan Pusan Statistik Provinsi Bali dilakukan beberapa metode yang digunakan seperti pada Gambar 2.1. Secara garis besar, pelaksanaan ini terdiri dari 5 tahap, yaitu pengumpulan dataset, split dataset, text processing, tahap klasifikasi, dan tahap integrasi website. Dataset yang digunakan dalam pelaksanaan ini berasal dari data primer internal hasil sruvei pencacah oleh Badan Pusat Statistik Provinsi Bali KBLI 2013. Dataset berupa teks yang dikonversi ke format .xlsx. Dataset terdiri dari 394524 data dengan 5 kolom informasi dengan 16 label yang mewakili kode kategori 1 digit. Selanjutnya dataset disajikan dalam perbandingan 70:30 dari data training, dan data testing.



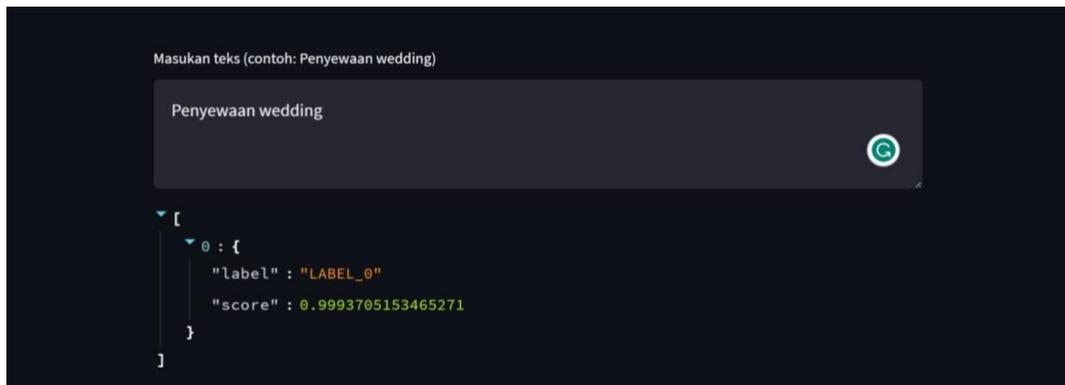
Gambar 2.1 Kerangka Pelaksanaan

Data yang sudah dibagi akan diproses terlebih dahulu untuk memperoleh data yang bersih dan siap digunakan melalui text processing. Selanjutnya barulah model dibangun dan dilatih menggunakan data yang telah bersih.

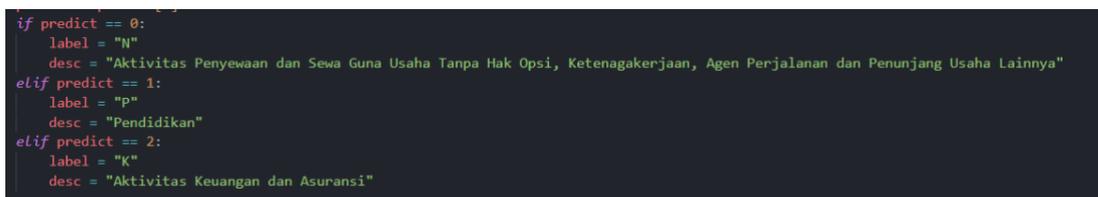
Setelah memperoleh model yang siap digunakan untuk mengimplemtasikan model dibangun sistem informasi dalam bentuk website.

3. HASIL DAN PEMBAHASAN

Gambar 3.1 dan Gambar 3.2 menunjukkan hasil tangkap layer sistem dari ASICA dalam bentuk website yang dibuat untuk mengimplementasikan dan uji coba sistem klasifikasi industri dan label yang berhasil untuk diklasifikasikan.



Gambar 3.1 Website Sistem Klasifikasi ASICA (Advandce Standart Industrial Classification Assistance)



Gambar 3.2 Baris Code Label Klasifikasi

Berdasarkan hasil tersbut dapat diketahui bahwa sistem telah berhasil mengklasifikasikan sesuai dengan kategori yang dimaksud yaitu penyewaan dan sewa guna usaha tanpa hal milik opsi dengan label 0 yang merepresentasikan kategori kode N.

4. KESIMPULAN DAN SARAN

Sistem klasifikasi industri sudah berhasil dibuat dan diimplementasikan dalam bentuk website dan diberikan kepada Badan Pusat Statistik Provinsi Bali. Sistem diterima dengan antusias karena proses klasifikasi dapat dibantu melalui sistem tersebut sebagai asisten sehingga mengurangi waktu pengerjaan nantinya. Namun, setiap sistem cerdas perlu tetap dikembangkan agar dapat menyelesaikan lebih banyak permasalahan, dalam hal ini sistem ini baru hanya dapat menyelesaikan klasifikasi kategori 1-digit yang terdiri atas 3 label, dikarenakan kurangnya dataset yang bersih, sehingga perlu dilakukan pendataan yang lebih teratur untuk meningkatkan performa dari sistem ini.

UCAPAN TERIMA KASIH

Dengan selesainya pelaksanaan ini, bukanlah menjadi sebuah akhir, melainkan suatu awal yang baru untuk memulai pengembangan lebih lanjut. Penulis menyadari betul bahwa ada orang-orang yang berjasa dibalik selesainya pelaksanaan ini. Tidak ada persembahan terbaik yang dapat penulis berikan selain rasa ucapan

terimakasih kepada pihak yang telah banyak membantu penulis. Secara khusus, penulis mengucapkan terimakasih kepada Agus Muliantara, S.Kom., M.Kom selaku dosen pembimbing yang telah sabar, meluangkan waktu, merelakan tenaga dan pikiran serta turut memberi perhatian dalam memberikan pendampingan selama proses penulisan skripsi ini. Terimakasih juga kepada lembaga Badan Pusat Statistik Provinsi Bali atas diberikannya kesempatan untuk dapat melakukan pelaksanaan pengabdian di sana. Segala kekurangan dan ketidaksempurnaan pelaksanaan ini, penulis sangat mengharapkan masukan, krtikan, dan saran yang bersifat membangun kearah perbaikan. Cukup banyak kesulitan yang penulis alami dalam penyusunan pengembangan sistem ini, tetapi Puji Tuhan dapat terselesaikan dengan baik. Akhir kata, penulis berharap semoga sistem ini dapat bermanfaat bagi semua pihak.

DAFTAR PUSTAKA

- BPS. (2023). Perusahaan Industri Pengolahan. *Badan Pusat Statistik*.
- Juarto, B., & Yulianto. (2023). Indonesian News Classification Using IndoBert. *International Journal of Intelligent Systems and Applications in Engineering*, 11(2).
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Saadah, S., Kaenova Mahendra Auditama, Ananda Affan Fattahila, Fendi Irfan Amorokhman, Annisa Aditsania, & Aniq Atiqi Rohmawati. (2022). Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(4). <https://doi.org/10.29207/resti.v6i4.4215>
- Sidik, A. D. W. M., Himawan Kusumah, I., Suryana, A., Edwinanto, Artiyasa, M., & Pradiftha Junfithrana, A. (2020). Gambaran Umum Metode Klasifikasi Data Mining. *FIDELITY: Jurnal Teknik Elektro*, 2(2). <https://doi.org/10.52005/fidelity.v2i2.111>
- Zong, C., Xia, R., & Zhang, J. (2021). Text Data Mining. In *Text Data Mining*. <https://doi.org/10.1007/978-981-16-0100-2>