

	<i>tradisional.</i>	
--	---------------------	--

Tabel 8. Hasil Terjemahan Bahasa Indonesia ke Bahasa Kaili

Bahasa Indonesia	Bahasa Kaili	Bahasa Kaili Hasil Model
demii apa pun tes cpns buat macet, sialan.	<i>Hurangah tes cpns nombiki macet, nadoyo.</i>	<i>I demii tes cpns namala nompaka macet, nambongo.</i>
Saya suka dengan suasananya yang nyaman untuk makan bersama keluarga. Di hari hari tertentu ada hiburan live musik. Makanannya pun punya pilihan yang beragam dari yang tradisional.	<i>Yaku nipokonoku ante suasananya nyaman bo mangande ante sarara. Eo eo tertenty naria hibura live music. Kandeana nadea pira dako ri tradisional.</i>	<i>Nipokonoku ante suasanana nanyama ka mangande ante sarara. Ri eo tertentu naria hiburan live music. Kandeana nadea dako tradisional.</i>

Dalam Tabel 7, dapat diamati bahwa kinerja model dalam mempelajari dan menghasilkan konteks terjemahan dari bahasa Indonesia ke bahasa Kaili cukup akurat. Hal ini karena ada beberapa kata dalam bahasa Kaili yang memiliki penggunaan yang serupa, seperti "nadoyo" dan "nambongo" (sebuah istilah dalam bahasa Kaili yang berarti "bodoh" atau "tolol"). Namun, ketika dihitung menggunakan SacreBLEU, yang lebih fokus pada pencocokan kata per kata, aspek ini mungkin tidak tercaptur, dan kinerjanya mungkin terlihat kurang baik. Situasi yang sama berlaku untuk model yang mempelajari terjemahan dari bahasa Kaili ke bahasa Indonesia. Makna kalimat terjemahan yang dihasilkan oleh model sebenarnya akurat, seperti yang terlihat dalam Tabel 8.

5. Kesimpulan

Hasil perbaikan dari pengumpulan data, pra-pemrosesan, dan penyetelan hiperparameter untuk model IndoBART-v2 telah berhasil menerjemahkan kalimat dari bahasa Indonesia ke bahasa Kaili dan dari bahasa Kaili ke bahasa Indonesia. Dengan menggunakan pra-pemrosesan tambahan, yaitu standardisasi kata menggunakan "Kamus Leksikon Bahasa Indonesia Keseharian", kinerja meningkat sebesar 38,7%. Dari eksperimen hiperparameter dengan menyesuaikan tingkat pembelajaran (learning rate) dan gamma, ditemukan bahwa parameter default adalah yang terbaik dengan tingkat pembelajaran 1e-4 dan gamma 0,9.

Skor kinerja akhir model IndoBART-v2, yang dinilai menggunakan metrik SacreBLEU, mencakup evaluasi dua skenario. Skenario 1 melibatkan penggunaan dataset yang mencakup ulasan bahasa formal dan informal dengan topik acak sebagai data sumber untuk pengembangan model, dan ulasan bahasa formal sebagai dataset uji. Skenario 2, di sisi lain, fokus pada penggunaan ulasan bahasa formal saja untuk pengembangan model dan dataset uji, khususnya dalam konteks terjemahan Indonesia-Kaili. Menariknya, skenario 1 menunjukkan efektivitas yang jauh lebih tinggi, dengan perbedaan yang signifikan sebesar 1,4. Begitu pula, dalam skema terjemahan Kaili-Indonesia, skenario 1 lebih unggul dibandingkan skenario 2, dengan perbedaan sebesar 0,3. Temuan ini menunjukkan bahwa menggabungkan topik-topik yang beragam, pola bahasa formal, dan gaya penulisan yang berbeda tidak memiliki dampak signifikan pada kinerja terjemahan dalam dataset uji jenis kalimat yang ditinjau.

Namun, skor SacreBLEU belum mencapai harapan kinerja sebenarnya dari model IndoBART-v2. Meskipun model ini telah mencapai makna kontekstual dan akurasi bahasa dalam terjemahan kalimat, terdapat kasus di mana model menggunakan kata-kata sinonim yang memiliki penggunaan kata yang serupa. Aspek ini tidak diperhitungkan oleh metrik SacreBLEU. Sebagai saran untuk penelitian mendatang, disarankan untuk menjelajahi model-model alternatif, menggabungkan topik-topik spesifik ke dalam dataset pelatihan dan pengujian, serta melakukan evaluasi manusia terhadap hasil terjemahan.

Daftar Pustaka

[1] N. W. Putri, "Pergeseran bahasa daerah Lampung pada masyarakat kota Bandar Lampung," *Jurnal Penelitian Humaniora*, vol. 19, no. 2, pp. 77–86, 2018.

[2] S. Fatimah, "Kekerabatan bahasa Kulawi dan bahasa Kaili di Sulawesi Tengah," *Kandai*, vol. 13, no. 2, pp. 249–262, 2017.

[3] U. Yaser and E. Efendi, "Ungkapan bahasa kaili dialek rai di kecamatan sirenja kabupaten donggala," *Bahasa dan sastra*, vol. 3, no. 1, 2018.

[4] I. Lampe and H. B. Anriani, "Stereotipe, Prasangka dan Dinamika Antaretnik," *Jurnal Penelitian Pers dan Komunikasi Pembangunan*, vol. 20, no. 1, pp. 19–32, 2016.

[5] Anonim, "Ancaman Kepunahan Merebak Masif. Pusat Riset Kependudukan BRIN," 2016. <https://kependudukan.brin.go.id/liputan-media/ancamankepunahan-merebak-masif/>

- [6] G. I. Winata *et al.*, “NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages,” *arXiv preprint arXiv:2205.15960*, 2022.
- [7] Z. Abidin, A. Sucipto, and A. Budiman, “Penerjemahan Kalimat Bahasa Lampung-Indonesia Dengan Pendekatan Neural Machine Translation Berbasis Attention Translation of Sentence LampungIndonesian Languages With Neural Machine Translation Attention Based,” *J. Kelitbang*, vol. 6, no. 02, pp. 191–206, 2018.
- [8] W. Wongso, H. Lucky, and D. Suhartono, “Pre-trained transformer-based language models for Sundanese,” *J Big Data*, vol. 9, no. 1, pp. 1–17, 2022.
- [9] Y. Liu *et al.*, “Multilingual denoising pre-training for neural machine translation,” *Trans Assoc Comput Linguist*, vol. 8, pp. 726–742, 2020.
- [10] A. Setiawan, H. Sujaini, and A. B. P. Negara, “Implementasi Optical Character Recognition (OCR) pada Mesin Penerjemah Bahasa Indonesia ke Bahasa Inggris,” *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 135–141, 2017.
- [11] P. Bhattacharyya, *Machine translation*. CRC Press, 2015.
- [12] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi, and S. Jain, “Machine translation using deep learning: An overview,” in *2017 international conference on computer, communications and electronics (comtelix)*, IEEE, 2017, pp. 162–167.
- [13] G. E. Hinton, “How neural networks learn from experience,” *Sci Am*, vol. 267, no. 3, pp. 144–151, 1992.
- [14] A. Vaswani *et al.*, “Attention is all you need,” *Adv Neural Inf Process Syst*, vol. 30, 2017.
- [15] S. Cahyawijaya *et al.*, “IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation,” *arXiv preprint arXiv:2104.08200*, 2021.
- [16] B. Wilie *et al.*, “IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding,” *arXiv preprint arXiv:2009.05387*, 2020.
- [17] G. Wenzek *et al.*, “CCNet: Extracting high quality monolingual datasets from web crawl data,” *arXiv preprint arXiv:1911.00359*, 2019.
- [18] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.
- [19] M. Post, “A call for clarity in reporting BLEU scores,” *arXiv preprint arXiv:1804.08771*, 2018.
- N. A. Salsabila, Y. A. Winatmoko, A. A. Septiandri, and A. Jamal, “Colloquial indonesian lexicon,” in *2018 International Conference on Asian Language Processing (IALP)*, IEEE, 2018, pp. 226–229.

Lampiran

Lampiran dapat berupa detil data dan contoh lebih lengkapnya, data-data pendukung, detail hasil pengujian, analisis hasil pengujian, detail hasil survey, surat pernyataan dari tempat studi kasus, screenshot tampilan sistem, hasil kuesioner dan lain-lain.