

ABSTRAK

Optical Character Recognition (OCR) berbasis *Large Multimodal Model* (LMM) menawarkan paradigma baru untuk ekstraksi data terstruktur dari dokumen resmi. Penelitian ini melakukan evaluasi komparatif performa dua model Google Gemini 2.5, yaitu Gemini 2.5 Pro dan Gemini 2.5 Flash, dalam tugas OCR untuk ekstraksi data terstruktur Kartu Keluarga (KK) Indonesia pada konteks aplikasi SmartGovt milik PT Puskomedia Indonesia Kreatif (Puskomedia). Untuk memenuhi prinsip perlindungan data pribadi sebagaimana diamanatkan UU Nomor 23 Tahun 2006 tentang Administrasi Kependudukan dan UU Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi, *dataset* disusun secara sintesis: 243 *record* KK fiktif di-generate dalam format JSON, dirender ke *template* HTML KK Indonesia melalui Playwright, kemudian diperluas dengan tujuh augmentasi visual (rotasi, perspektif, gelap, terang, *blur*, kompresi JPEG, dan resolusi rendah) sehingga total *dataset* evaluasi mencapai 1.944 gambar. Variabel penelitian mencakup 9 *field header* dan 14 *field* anggota per individu. Metrik evaluasi yang digunakan adalah *Exact Match Accuracy*, *Character Error Rate* (CER), *Word Error Rate* (WER), serta akurasi *field* kritis NIK 16 digit dan Nomor KK. Uji signifikansi menggunakan Wilcoxon *Signed-Rank Test* pada taraf $\alpha=0,05$ dengan *effect size rank-biserial correlation*. Hasil penelitian menunjukkan bahwa Gemini 2.5 Pro mencapai *Exact Match Accuracy* 99,90%, CER 0,03%, dan WER 0,09%, sedangkan Gemini 2.5 Flash hanya mencapai 56,91%, 39,58%, dan 46,14%. Akurasi NIK Pro mencapai 100,00% sedangkan Flash hanya 46,90%. Uji Wilcoxon menghasilkan $p < 0,001$ dengan *rank-biserial* $r = -0,9994$ (*effect size* sangat besar). Analisis lebih lanjut mengungkap bahwa Gemini 2.5 Flash menunjukkan *failure mode* khas berupa fabrikasi data pada gambar bersih, *anchoring prior knowledge* geografis, dan ketidakstabilan jumlah anggota keluarga (halusinasi dan omisi). Kesimpulan: Gemini 2.5 Pro layak direkomendasikan sebagai model utama untuk implementasi OCR KK pada SmartGovt, sedangkan pada *dataset* sintesis terkontrol ini Gemini 2.5 Flash menunjukkan ketidaksiapan substansial; generalisasi ke KK fisik memerlukan validasi lanjutan.

Kata kunci: OCR, Kartu Keluarga, Gemini 2.5 Pro, Gemini 2.5 Flash, *Large Multimodal Model*, CER, WER, evaluasi komparatif, *dataset* sintesis, halusinasi.

ABSTRACT

Optical Character Recognition (OCR) based on Large Multimodal Models (LMM) offers a new paradigm for structured data extraction from official documents. This research conducts a comparative evaluation of two Google Gemini 2.5 models, namely Gemini 2.5 Pro and Gemini 2.5 Flash, on the task of structured data extraction from Indonesian Family Cards (Kartu Keluarga / KK) in the context of the SmartGovt application by PT Puskomedia Indonesia Kreatif (Puskomedia). To comply with personal-data-protection principles mandated by Indonesian Law No. 23/2006 on Civil Administration and Law No. 27/2022 on Personal Data Protection, the dataset was constructed synthetically: 243 fictitious KK records were generated in JSON format, rendered to an Indonesian KK HTML template via Playwright, and expanded with seven visual augmentations (rotation, perspective, dark, bright, blur, JPEG compression, and low resolution) yielding a total of 1,944 evaluation images. The research variables include 9 header fields and 14 member fields per individual. Evaluation metrics used are Exact Match Accuracy, Character Error Rate (CER), Word Error Rate (WER), and accuracy of critical fields (16-digit NIK and KK Number). Statistical significance was assessed using the Wilcoxon Signed-Rank Test at $\alpha=0.05$ along with rank-biserial correlation effect size. Results show that Gemini 2.5 Pro achieves 99.90% Exact Match, 0.03% CER, and 0.09% WER, whereas Gemini 2.5 Flash only achieves 56.91%, 39.58%, and 46.14% respectively. NIK accuracy reaches 100.00% for Pro versus 46.90% for Flash. The Wilcoxon test yields $p < 0.001$ with rank-biserial $r = -0.9994$ (very large effect size). Further analysis reveals that Gemini 2.5 Flash exhibits distinctive failure modes: fabrication of data on clean images, geographic prior-knowledge anchoring, and instability in the number of family members (hallucination and omission). The study concludes that Gemini 2.5 Pro is recommended as the primary model for KK OCR implementation in SmartGovt, whereas on this controlled synthetic dataset Gemini 2.5 Flash shows substantial unreadiness; generalization to physical KK requires further validation.

Keywords: OCR, Family Card, Gemini 2.5 Pro, Gemini 2.5 Flash, Large Multimodal Model, CER, WER, comparative evaluation, synthetic dataset, hallucination.