

# Pipeline Ekstraksi Adaptif untuk Otomatisasi Administrasi Kependudukan Sistem Informasi Desa

Reza Ikhsanda<sup>1\*</sup>, Wahyu Nugraha<sup>2</sup>, Rabiatus Saadah<sup>3</sup>

<sup>1,2,3</sup>Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: <sup>1\*</sup>15220712@bsi.ac.id, <sup>2</sup>wahyu.whn@bsi.ac.id, <sup>3</sup>rabiatus.rbh@bsi.ac.id

Received: 25 Juni 2026 | Revision: 30 Juni 2026 | Accepted: 1 Juli 2026

## Abstrak

Administrasi kependudukan di tingkat desa masih banyak bergantung pada pengetikan ulang data Kartu Keluarga secara manual sehingga lambat dan rentan kesalahan. Penelitian ini bertujuan mengembangkan PEDEs, sistem informasi desa berbasis web yang mengotomatisasi ekstraksi data Kartu Keluarga dari berkas PDF melalui pipeline ekstraksi adaptif. Sistem menganalisis struktur internal berkas PDF, lalu memilih jalur teks bila tersedia lapisan teks digital atau jalur visual dengan pra-pemrosesan citra bila berkas hanya berupa hasil pemindaian, kemudian melakukan ekstraksi dokumen semantik untuk menghasilkan data terstruktur berformat JSON. Stabilitas proses dijaga melalui sistem antrian dan pembatasan laju, sedangkan keakuratan dijaga melalui sanitasi otomatis dan verifikasi operator. Sistem dibangun menggunakan Laravel, Inertia.js, dan React serta memanfaatkan layanan Groq sebagai mesin ekstraksi semantik. Pengujian fungsional black-box terhadap 24 skenario seluruhnya berstatus sesuai. Pengujian akurasi pada 17 dokumen menghasilkan precision 92,00% dan F1-score rata-rata 85,44%, dengan 98,59% pada PDF digital dan 96,38% pada PDF hasil pemindaian. Dibandingkan proses manual, sistem memangkas waktu pemrosesan dari 120 detik menjadi 10 detik per dokumen (91,67%). Hasil ini menunjukkan pipeline ekstraksi adaptif mampu mempercepat administrasi kependudukan desa sekaligus menekan human error.

**Kata Kunci:** administrasi kependudukan, ekstraksi dokumen semantik, pipeline ekstraksi adaptif, pra-pemrosesan citra, sistem informasi desa

## Abstract

Population administration at the village level still largely depends on manually retyping Family Card (Kartu Keluarga) data, making it slow and error-prone. This study aims to develop PEDEs, a web-based village information system that automates the extraction of Family Card data from PDF files through an adaptive extraction pipeline. The system analyzes the internal structure of each PDF file, then selects a text path when a digital text layer is available or a visual path with image preprocessing when the file is only a scanned image, and subsequently performs semantic document extraction to produce structured data in JSON format. Process stability is maintained through a queue system and rate limiting, while accuracy is preserved through automatic sanitization and operator verification. The system was built using Laravel, Inertia.js, and React, and leverages the Groq service as the semantic extraction engine. Black-box functional testing of 24 scenarios all passed. Accuracy testing on 17 documents yielded 92.00% precision and an average F1-score of 85.44%, with 98.59% on digital PDFs and 96.38% on scanned PDFs. Compared with the manual process, the system cut processing time from 120 to 10 seconds per document (91.67%). These results show that the adaptive extraction pipeline can accelerate village population administration while reducing human error.

**Keywords:** adaptive extraction pipeline, image preprocessing, population administration, semantic document extraction, village information system

## 1. PENDAHULUAN

Transformasi digital sektor pemerintahan telah menjadi agenda nasional. Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik (SPBE) menempatkan teknologi informasi sebagai instrumen utama peningkatan kualitas dan akuntabilitas pelayanan publik hingga ke tingkat pemerintahan desa [1]. Indonesia memiliki lebih dari tujuh puluh empat ribu desa yang dituntut menyelenggarakan tata kelola administrasi yang tertib dan dapat dipertanggungjawabkan sesuai Undang-Undang Nomor 6 Tahun 2014 tentang Desa [2], sehingga modernisasi pengelolaan data menjadi prasyarat agar desa mampu menjalankan fungsi pelayanannya secara optimal di tengah arus digitalisasi yang kian menguat [3].

Administrasi kependudukan menempati posisi sentral karena hampir seluruh layanan desa bertumpu pada ketersediaan dan keakuratan data penduduk. Perangkat desa memanfaatkan data kependudukan sebagai basis penerbitan surat keterangan, pendataan penerima bantuan sosial, rekapitulasi jumlah warga, hingga pelaporan administratif ke tingkat yang lebih tinggi. Berbagai sistem informasi desa telah dikembangkan untuk mendukung pelayanan ini [4], [5], [6], namun pengelolaan data kependudukan di sebagian besar desa masih bertumpu pada dokumen fisik Kartu Keluarga. Perangkat desa membaca lembaran Kartu Keluarga lalu mengetik ulang setiap atribut—nomor Kartu Keluarga, Nomor Induk Kependudukan (NIK), nama, alamat, tanggal lahir, pendidikan, dan pekerjaan—ke dalam berkas Microsoft Excel. Rutinitas ini berlangsung lambat, berulang, dan rentan kesalahan: digit NIK tertukar atau nama salah baca dapat menjalar ke seluruh layanan turunan seperti penerbitan surat keterangan dan penetapan penerima bantuan sosial. Berbagai kajian

menegaskan bahwa pengelolaan data secara manual merupakan hambatan struktural yang memengaruhi efisiensi kerja sekaligus validitas data administrasi [3].

Sebagai langkah awal digitalisasi, banyak desa memindai Kartu Keluarga dan menyimpannya dalam format Portable Document Format (PDF). Namun berkas yang dihasilkan tidak memiliki struktur internal yang seragam: sebagian menyimpan lapisan teks digital (*text layer*) yang dapat dibaca langsung oleh mesin karena perangkat pemindai tertentu menjalankan OCR otomatis, sedangkan sebagian lain hanya berisi citra hasil pemindaian tanpa lapisan teks. Perbedaan karakteristik internal ini menyebabkan satu metode pembacaan tidak dapat diterapkan secara seragam. Kajian mengenai ekstraksi informasi dari dokumen PDF menyimpulkan bahwa keragaman struktur ini menuntut pendekatan pemrosesan yang adaptif, yaitu kemampuan menyesuaikan strategi ekstraksi berdasarkan karakteristik internal setiap dokumen, bukan satu metode tunggal [7].

Pendekatan *Optical Character Recognition* (OCR) konvensional yang lazim digunakan untuk mengubah citra menjadi teks memiliki keterbatasan ketika dihadapkan pada dokumen administrasi kependudukan, karena hanya berfokus pada pengenalan karakter tanpa memahami struktur tabel maupun keterkaitan *semantic* antaratribut [8]. Padahal Kartu Keluarga merupakan dokumen semi-terstruktur yang menuntut sistem memahami posisi kolom serta hubungan antaranggota keluarga. Pendekatan *intelligent document processing* dan ekstraksi dokumen *semantic* terbukti meningkatkan kualitas pemahaman dokumen semi-terstruktur dibandingkan OCR tradisional [9].

Pada dokumen hasil pemindaian, kualitas citra menjadi penentu keberhasilan ekstraksi sehingga tahap pra-pemrosesan citra (*image preprocessing*)—koreksi orientasi, konversi keabuan (*grayscale*), peningkatan kontras, dan penajaman—diperlukan untuk memperbaiki keterbacaan masukan [10]. Selain ketepatan, stabilitas proses tidak dapat diabaikan: pemrosesan seluruh dokumen secara serentak tanpa kendali berisiko menimbulkan lonjakan beban, sehingga diperlukan *queue system* dan *rate limiting* agar pekerjaan ekstraksi berjalan terkendali [11]. Hasil ekstraksi otomatis pun sebaiknya tidak langsung dianggap final, melainkan diverifikasi operator sebagai kontrol kualitas sebelum masuk ke basis data utama.

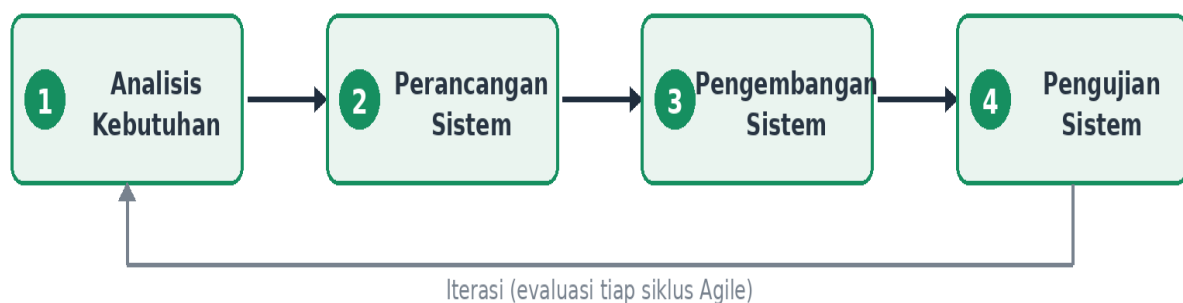
Penelitian terdahulu cenderung membahas aspek-aspek tersebut secara terpisah. Burhanuddin et al. mengembangkan sistem informasi kependudukan desa berbasis QR Code namun berfokus pada pengamanan data, bukan ekstraksi dokumen [12]. Atagong et al. menelaah teknik ekstraksi PDF tetapi berhenti pada tataran tinjauan [7], sementara Islam et al. menunjukkan keunggulan ekstraksi *semantic* namun belum mengintegrasikannya ke dalam sistem informasi desa yang utuh [9]. Pada ranah restorasi citra, Guan et al. [10] dan Liu [13] memadukan pra-pemrosesan citra dan *deep learning* dengan OCR, namun belum menyoar dokumen kependudukan dengan kebutuhan pemahaman relasi antaratribut. Maharjan et al. menegaskan peran antrian dalam menjaga stabilitas sistem berbeban tinggi, tetapi tidak dikaitkan dengan ekstraksi dokumen [11].

Berdasarkan telaah tersebut, belum ditemukan penelitian yang menyatukan deteksi struktur PDF adaptif, ekstraksi dokumen *semantic*, pra-pemrosesan citra, manajemen antrian, dan verifikasi operator dalam satu kesatuan sistem informasi desa. Penelitian ini mengisi celah tersebut dengan mengembangkan PEDES (Pengelolaan Data Desa), sistem informasi desa berbasis web yang menerapkan *pipeline* ekstraksi adaptif. Tujuan penelitian adalah merancang dan membangun sistem tersebut serta menguji akurasi ekstraksi dan efisiensinya dibandingkan proses manual. Artikel ini diorganisasikan ke dalam bagian metodologi penelitian, hasil dan pembahasan, serta kesimpulan.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menggunakan metode pengembangan perangkat lunak *Agile* yang bersifat iteratif dan fleksibel terhadap hasil evaluasi pada setiap iterasi [14], [15], [16]. Pengumpulan data dilakukan melalui observasi proses administrasi kependudukan di Kantor Desa Mega Timur, wawancara dengan perangkat desa, dan studi pustaka. Tahapan penelitian terdiri atas empat fase yang dijalankan secara berulang, sebagaimana ditunjukkan pada Gambar 1.



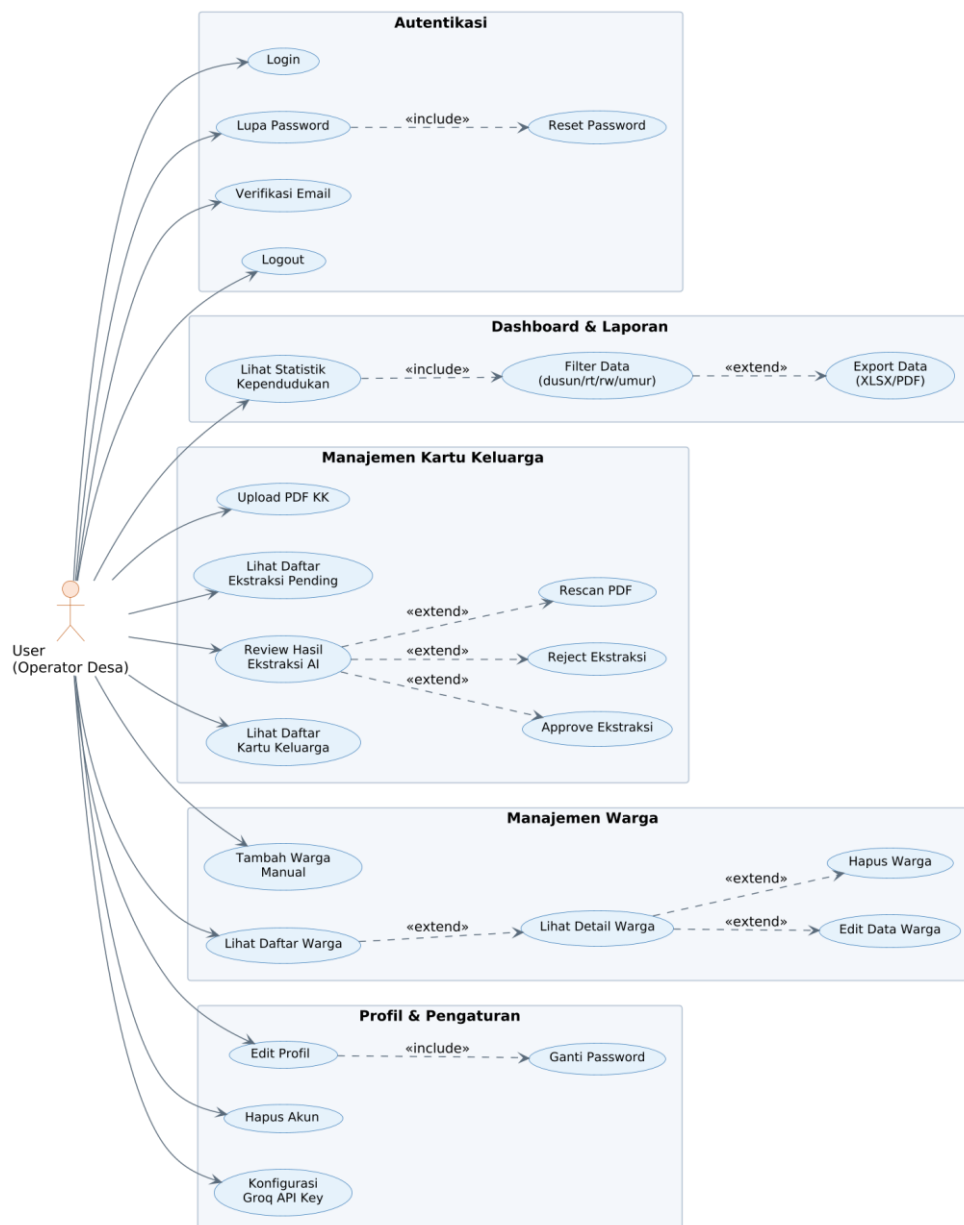
**Gambar 1.** Tahapan penelitian dengan pendekatan Agile

Tahap analisis kebutuhan mengidentifikasi proses administrasi kependudukan, struktur data Kartu Keluarga standar nasional, serta permasalahan rekapitulasi data warga yang masih dilakukan secara manual. Tahap perancangan sistem

menghasilkan arsitektur sistem, struktur basis data, alur *pipeline* ekstraksi adaptif, dan rancangan antarmuka. Tahap pengembangan membangun seluruh fungsi aplikasi, sedangkan tahap pengujian mengukur akurasi ekstraksi serta membandingkan efisiensinya terhadap proses manual. Hasil evaluasi pada setiap iterasi digunakan untuk menyempurnakan rancangan pada iterasi berikutnya.

## 2.2 Pemodelan Kebutuhan dengan UML

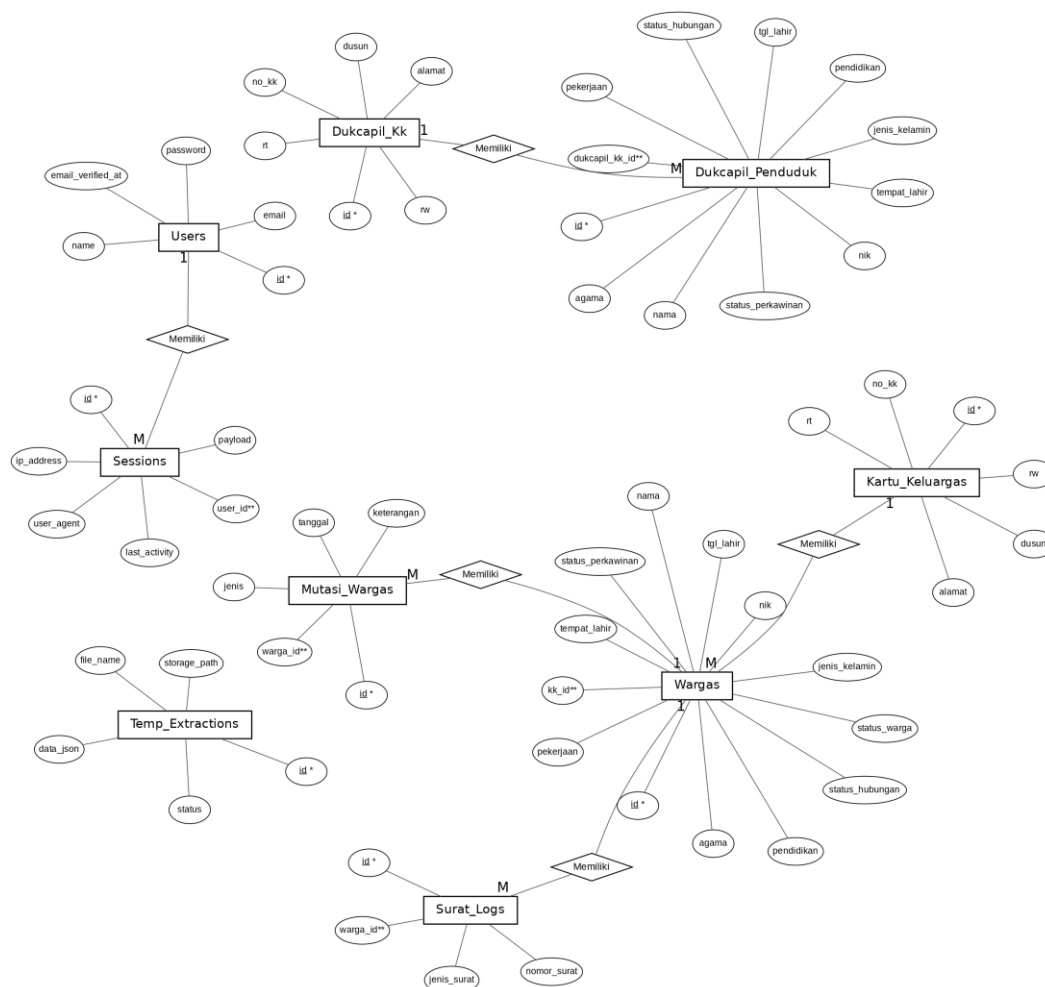
Kebutuhan fungsional dimodelkan menggunakan *use case diagram* (*Unified Modeling Language/UML*) [17] yang menggambarkan aktor tunggal, yaitu operator desa terautentikasi, beserta seluruh fungsi yang dapat diaksesnya: autentikasi, unggah dan ekstraksi Kartu Keluarga, review dan verifikasi, manajemen warga, dashboard dan ekspor data, serta pengaturan, sebagaimana disajikan pada Gambar 2. Struktur perangkat lunak dirancang berlapis—lapisan *Model*, *Controller*, layanan domain (*Service*, antara lain *KkExtractorService*), dan pekerjaan antrian (*Job*)—untuk memisahkan tanggung jawab serta memudahkan pemeliharaan dan pengembangan fitur lanjutan.



**Gambar 2.** Use case diagram aplikasi PEDES

## 2.3 Perancangan Basis Data

Basis data dirancang relasional dengan dua entitas utama, yaitu *kartu\_keluargas* dan *wargas*, yang berelasi satu-ke-banyak sehingga satu Kartu Keluarga menaungi banyak anggota. Entitas *temp\_extractions* menampung hasil ekstraksi sementara sebelum diverifikasi, sedangkan entitas *pembanding Dukcapil* digunakan untuk sinkronisasi data. Rancangan *Entity Relationship Diagram* (ERD) basis data disajikan pada Gambar 3, mengikuti kaidah normalisasi agar data tersimpan konsisten dan minim redundansi [18].



**Gambar 3.** Entity Relationship Diagram (ERD) basis data PEDES

### 2.4 Arsitektur Sistem dan Pipeline Ekstraksi Adaptif

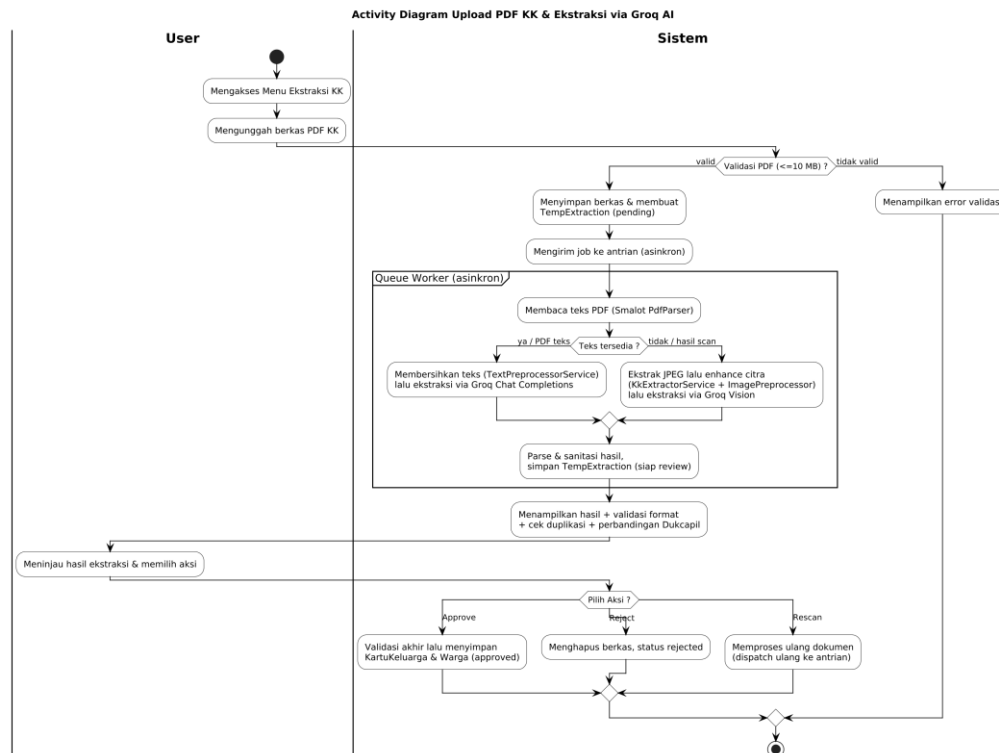
Sistem PEDES dibangun sebagai aplikasi berbasis web [19] dengan mengacu pada konsep sistem informasi [20], [21], menggunakan kerangka kerja (*framework*) Laravel 13 pada sisi *back-end* dengan PHP 8.3, antarmuka berbasis Inertia.js dan React, serta basis data MySQL. Pembacaan teks PDF memanfaatkan pustaka Smalot PdfParser, sedangkan ekstraksi dokumen *semantic* memanfaatkan layanan Groq (model llama-3.3-70b-versatile) yang mengembalikan data terstruktur berformat JSON. Stabilitas proses dikelola melalui Laravel Queue dengan *driver* basis data, satu *worker*, pembatasan laju lima belas permintaan per menit, pengulangan otomatis, dan batas waktu (*timeout*) sembilan puluh detik. Ringkasan lingkungan dan komponen teknologi yang digunakan disajikan pada Tabel 1.

**Tabel 1.** Spesifikasi Teknis dan Lingkungan Implementasi

Komponen	Spesifikasi
Bahasa & framework back-end	PHP 8.3, Laravel 13
Front-end	Inertia.js, React, Tailwind CSS
Basis data	MySQL
Pustaka parsing PDF	Smalot PdfParser
Pengolah citra	Ekstraksi & penyempurnaan JPEG tertanam (rotate, upscale, grayscale, kontras, sharpen)
Mesin ekstraksi semantik	Groq API (Chat Completions & Vision), model llama-3.3-70b-versatile
Antrian (queue)	Laravel Queue, driver database
Web server	Apache
Perangkat keras (uji)	Intel Core i7 gen-12, RAM 16 GB, SSD

Sumber: Hasil penelitian (2026)

Prosedur inti sistem adalah *pipeline* ekstraksi adaptif yang menentukan jalur pemrosesan berdasarkan karakteristik internal berkas, sebagaimana diilustrasikan pada Gambar 4 dan dirangkum pada Tabel 2.



**Gambar 4.** Diagram aktivitas alur unggah dan ekstraksi Kartu Keluarga

**Tabel 2.** Tahapan Pipeline Ekstraksi Adaptif

Tahap	Proses	Keterangan
1	Unggah Dokumen	Operator mengunggah ( <i>upload</i> ) PDF Kartu Keluarga; sistem memvalidasi tipe dan ukuran berkas ( $\leq 10$ MB).
2	Data Sementara	Berkas disimpan dan dicatat sebagai data sementara berstatus <i>pending</i> .
3	Antrian	Tugas ekstraksi dimasukkan ( <i>dispatch</i> ) ke antrian dengan jeda yang dapat dikonfigurasi.
4	Analisis PDF	Sistem memeriksa keberadaan lapisan teks digital pada berkas.
5	Jalur Teks	Bila lapisan teks tersedia, teks dibersihkan lalu dikirim ke mesin ekstraksi semantik.
6	Jalur Visual	Bila tidak, JPEG tertanam diekstraksi dan disempurnakan ( <i>rotate, upscale, grayscale, kontras, sharpen</i> ).
7	Ekstraksi Semantik	Layanan Groq mengembalikan JSON berisi data Kartu Keluarga dan anggota keluarga.
8	Sanitasi Data	Format 16 digit nomor KK dan NIK divalidasi serta atribut dinormalisasi.
9	Verifikasi	Hasil dibandingkan dengan data Dukcapil dan diberi opsi <i>Approve/Reject/Rescan</i> .
10	Penyimpanan Final	Pada persetujuan, data ditulis ke basis data utama.

Sumber: Hasil penelitian (2026)

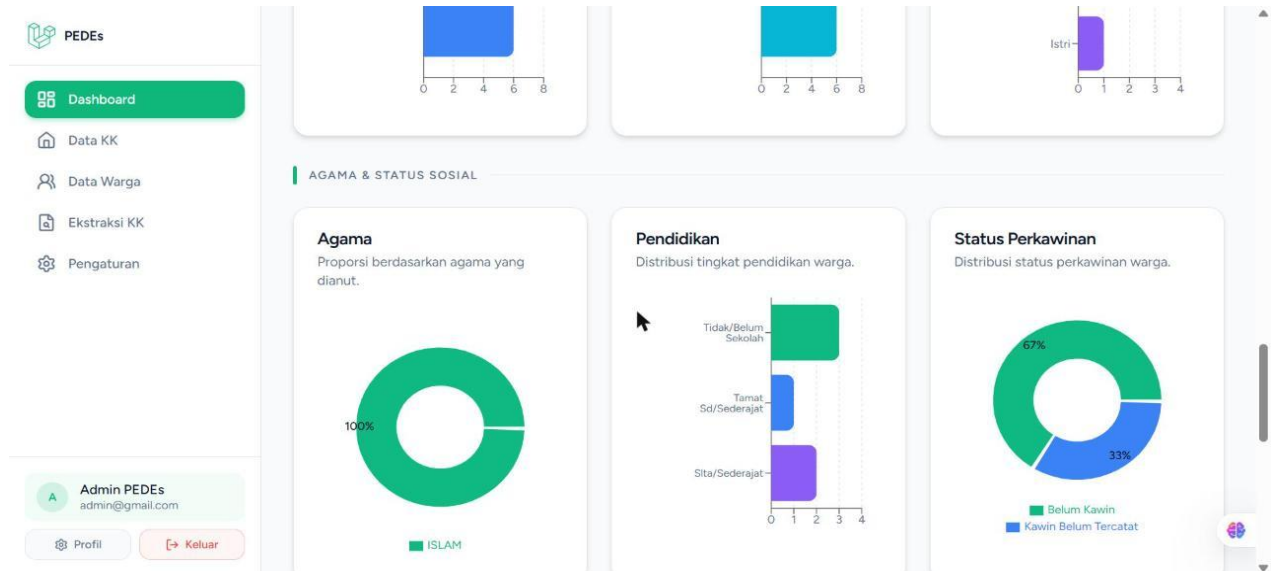
## 2.5 Metode Pengujian

Pengujian dilakukan dalam dua tahap. Pengujian fungsional menggunakan metode *black-box*, yaitu menguji kesesuaian masukan dan keluaran tiap fungsi tanpa memperhatikan struktur kode [22]. Pengujian akurasi membandingkan data hasil ekstraksi terhadap data sebenarnya pada dokumen Kartu Keluarga uji yang mencakup PDF digital, PDF hasil pemindaian, dokumen berkualitas rendah, dan dokumen posisi miring, dengan metrik akurasi atribut, *precision*, *recall*, dan *F1-score*. Efisiensi dinilai dengan membandingkan waktu pemrosesan otomatis terhadap input manual pada sejumlah dokumen yang sama.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Implementasi Antarmuka Sistem

Aplikasi PEDEs berhasil diimplementasikan sebagai aplikasi web yang responsif. Operator masuk melalui halaman login, kemudian diarahkan ke dashboard yang menampilkan kartu statistik kependudukan dan visualisasi demografi secara real-time (Gambar 5), sehingga kondisi data desa dapat dipantau dengan cepat.



**Gambar 5.** Dashboard statistik dan visualisasi demografi

Selain dashboard, sistem menyediakan halaman Data Kartu Keluarga dan Daftar Warga yang menampilkan seluruh data kependudukan lengkap dengan fitur pencarian dan penyaringan, halaman detail warga, formulir penambahan warga manual, serta halaman pengaturan untuk mengonfigurasi kunci layanan ekstraksi. Proses ekstraksi otomatis dilakukan pada halaman unggah dan antrian, tempat operator mengunggah berkas PDF Kartu Keluarga dan memantau status pemrosesan yang berjalan bertahap dan terkendali.

#### 3.2 Pengujian Fungsional

Pengujian fungsional *black-box* dilakukan langsung pada aplikasi yang berjalan melalui peramban. Sebanyak 24 skenario uji dievaluasi dan seluruhnya berstatus Sesuai, sebagaimana dirangkum pada Tabel 3. Hal ini membuktikan bahwa fitur inti sistem berfungsi sesuai rancangan.

**Tabel 3. Ringkasan Hasil Pengujian Fungsional (Black-box)**

No	Skenario Pengujian	Jumlah Kasus	Hasil
1	Autentikasi (login, logout, reset password)	3	Sesuai
2	Unggah berkas PDF Kartu Keluarga	2	Sesuai
3	Ekstraksi adaptif (jalur teks & visual)	3	Sesuai
4	Sanitasi data (NIK & nomor KK 16 digit)	2	Sesuai
5	Sinkronisasi data Dukcapil	2	Sesuai
6	Review & verifikasi (approve/reject/rescan)	3	Sesuai
7	Manajemen warga (tambah, ubah, hapus, lihat)	4	Sesuai
8	Manajemen Kartu Keluarga	2	Sesuai
9	Dashboard & visualisasi demografi	2	Sesuai
10	Ekspor data (Excel/PDF)	1	Sesuai
	<b>Total</b>	<b>24</b>	<b>Sesuai</b>

Sumber: Hasil penelitian (2026)

Secara khusus, *pipeline* ekstraksi adaptif berhasil memproses berkas PDF digital: lapisan teks terbaca oleh Smalot PdfParser, muatan dikirim ke model llama-3.3-70b-versatile, dan keluaran JSON mempertahankan relasi satu-ke-banyak antara satu Kartu Keluarga dan enam anggotanya. Hasil ekstraksi ditampilkan kepada operator untuk diverifikasi beserta peringatan validasi dan perbandingan dengan data Dukcapil sebagai kontrol kualitas tambahan (Gambar 6).

Temuan ini sejalan dengan Islam et al. bahwa ekstraksi *semantic* mampu mempertahankan struktur dokumen semi-terstruktur lebih baik daripada OCR tradisional [9].



**Gambar 6.** Tampilan review hasil ekstraksi beserta peringatan validasi

### 3.3 Akurasi Ekstraksi

Evaluasi akurasi terhadap 17 dokumen dalam empat kategori menunjukkan kinerja yang bervariasi sesuai karakteristik masukan, sebagaimana disajikan pada Tabel 4.

**Tabel 4.** Hasil Pengujian Akurasi Ekstraksi

Jenis Dokumen	Jumlah	Akurasi Atribut	Precision	Recall	F1-score
PDF digital	1	98,59%	98,59%	98,59%	98,59%
PDF hasil pemindaian	8	96,11%	96,65%	96,11%	96,38%
Dokumen kualitas rendah	3	65,26%	98,58%	65,26%	78,62%
Dokumen posisi miring	5	63,10%	74,17%	63,10%	68,18%
Rata-rata keseluruhan	17	80,77%	92,00%	80,77%	85,44%

Sumber: Hasil penelitian (2026)

Pada PDF digital, sistem mencapai *F1-score* 98,59%, menandakan jalur ekstraksi teks menghasilkan keluaran nyaris sempurna. Pada delapan PDF hasil pemindaian, *F1-score* 96,38% mengonfirmasi bahwa pra-pemrosesan citra berhasil meningkatkan keterbacaan sebelum ekstraksi *semantic*. Sebaliknya, dokumen berkualitas rendah mencatat *F1-score* 78,62% dengan *precision* tinggi (98,58%) namun *recall* rendah (65,26%), mengindikasikan sebagian atribut gagal diekstraksi akibat keterbatasan kualitas citra; dokumen posisi miring mencatat *F1-score* 68,18%, menunjukkan koreksi orientasi belum sepenuhnya memulihkan keterbacaan pada kemiringan ekstrem. Secara keseluruhan, rata-rata *F1-score* 85,44% dengan *precision* 92,00% mencerminkan sistem mampu mengekstraksi data secara akurat pada kondisi ideal dan tetap memberikan kinerja yang dapat diterima pada kondisi menantang.

### 3.4 Perbandingan Efisiensi

Perbandingan waktu pemrosesan manual dan otomatis disajikan pada Tabel 5.

**Tabel 5.** Perbandingan Efisiensi Proses Otomatis dan Manual

Aspek	Proses Manual	Proses Otomatis (PEDEs)	Selisih
Waktu per dokumen	120 detik	10 detik	110 detik (91,67%)
Waktu untuk 5 dokumen	600 detik	55 detik	545 detik
Potensi kesalahan input	Tinggi	Rendah	—

Sumber: Hasil penelitian (2026)

Proses manual membutuhkan rata-rata 120 detik per dokumen, sedangkan PEDEs menyelesaikannya dalam 10 detik—penghematan 110 detik atau 91,67% per dokumen. Pada simulasi lima dokumen, proses manual memerlukan 600 detik

sementara sistem otomatis hanya 55 detik; selisih ini berpotensi semakin membesar seiring bertambahnya volume dokumen. Selain memangkas waktu, sistem menekan potensi kesalahan input melalui sanitasi otomatis dan verifikasi operator, sehingga data yang masuk ke basis data utama melewati dua lapis pemeriksaan. Mekanisme antrian dengan *rate limit* lima belas *job* per menit dan satu *worker* terbukti menjaga stabilitas sistem tanpa kegagalan akibat lonjakan beban, selaras dengan temuan Maharjan et al. mengenai peran antrian pada sistem bervolume tinggi [11]. Sebagai kontrol kualitas tambahan, sinkronisasi Dukcapil berhasil mengidentifikasi dua perbedaan atribut status perkawinan antara hasil ekstraksi dan data pembandingan, sehingga operator dapat memutuskan pembaruan data.

#### 4. KESIMPULAN

Penelitian ini berhasil mengembangkan PEDEs, sistem informasi desa berbasis web dengan *pipeline* ekstraksi adaptif untuk mengotomatiskan ekstraksi data Kartu Keluarga dari berkas PDF digital maupun hasil pemindaian. Pengujian fungsional *black-box* terhadap 24 skenario seluruhnya berstatus Sesuai, dan sistem mencapai *precision* 92,00% serta *F1-score* rata-rata 85,44% pada 17 dokumen uji, dengan 98,59% pada PDF digital dan 96,38% pada PDF hasil pemindaian, serta memangkas waktu pemrosesan hingga 91,67% dibandingkan proses manual. Secara praktis, otomatisasi berbasis *pipeline* ekstraksi adaptif dapat mempercepat administrasi kependudukan desa sekaligus menekan human error sehingga mendukung tata kelola pemerintahan desa berbasis elektronik. Penelitian ini memiliki keterbatasan berupa penurunan kinerja pada dokumen berkualitas rendah dan posisi miring ekstrem serta pengujian yang masih terbatas pada satu desa, sehingga penelitian lanjutan disarankan menyempurnakan algoritma koreksi orientasi dan restorasi citra, memperluas dukungan ke jenis dokumen kependudukan lain, serta mengintegrasikan sistem dengan basis data kependudukan tingkat lebih tinggi dan mengujinya pada lebih banyak desa.

#### REFERENCES

- [1] Harmayni, E. N. Sihombing, dan O. Medaline, "Aspek hukum pemanfaatan Sistem Pemerintahan Berbasis Elektronik (SPBE) dalam penyelenggaraan administrasi negara," *Jurnal Ilmu Sosial & Hukum*, vol. 4, no. 3, pp. 3893–3902, 2026, doi: 10.61104/alz.v4i3.5848.
- [2] I. Sandika, S. Aini, Y. K. Simbolon, dan S. Hadiningrum, "Analisis sistem pemerintah desa di Indonesia," *Terang: Jurnal Kajian Ilmu Sosial, Politik dan Hukum*, vol. 1, no. 1, pp. 212–223, 2024, doi: 10.62383/terang.v1i1.89.
- [3] A. P. E. Susilowati, R. Rachmawati, dan R. Rijanta, "Smart village concept in Indonesia: ICT as determining factor," *Heliyon*, vol. 11, no. 1, p. e41657, 2025.
- [4] Y. M. Kristania, "Sistem informasi pelayanan administrasi kependudukan desa (M-Desa) dengan metode user centered design," *Indonesian Journal on Software Engineering*, vol. 7, no. 1, pp. 1–9, 2021.
- [5] W. G. S. Parwita, R. A. Mutiarani, dan I. N. W. Adnyana, "Implementasi aplikasi sistem informasi desa dan kependudukan berbasis web di Desa Kukuh Kerambitan," *Jurnal Widya Laksana*, vol. 10, no. 1, p. 27, 2021.
- [6] G. Z. Muflih dan R. Ardiansah, "Pengembangan sistem informasi administrasi kependudukan Desa Gumelem Wetan berbasis website," *Jurnal Kridatama Sains dan Teknologi*, vol. 5, no. 2, pp. 536–551, 2023.
- [7] S. D. Atagong et al., "A review on knowledge and information extraction from PDF documents and storage approaches," *Frontiers in Artificial Intelligence*, vol. 8, p. 1466092, 2025.
- [8] H. Holila, A. R. Pratama, S. A. P. Lestari, dan J. Indra, "Introduction national identification number and name on ID card using OCR method," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 4, pp. 1191–1196, 2024, doi: 10.52436/1.jutif.2024.5.4.2242.
- [9] M. M. Islam et al., "IDP accelerator: Agentic document intelligence from extraction to compliance validation," *arXiv preprint*, arXiv:2602.23481, 2026.
- [10] S. Guan et al., "PreP-OCR: A complete pipeline for document image restoration and enhanced OCR accuracy," in *Proc. 63rd Annu. Meeting of the Association for Computational Linguistics (ACL)*, 2025.
- [11] R. Maharjan, M. S. H. Chy, M. A. Arju, dan T. Cerny, "Benchmarking message queues," *Telecom*, vol. 4, no. 2, pp. 298–312, 2023.
- [12] A. Burhanuddin, F. Nugraha, D. L. Fithri, P. K. Handayani, dan N. Susanti, "Pemanfaatan teknologi QR Code untuk sistem informasi kependudukan di Desa Pilangrejo," *Jurnal SITECH: Sistem Informasi dan Teknologi*, vol. 6, no. 2, pp. 85–92, 2023.
- [13] Y.-T. Liu, "Printed document layout analysis and optical character recognition system based on deep learning," *Scientific Reports*, vol. 15, p. 23761, 2025.
- [14] R. S. Pressman dan B. R. Maxim, *Software Engineering: A Practitioner's Approach*, edisi ke-9. New York: McGraw-Hill, 2020.
- [15] S. H. Nova, A. P. Widodo, dan B. Warsito, "Analisis metode Agile pada pengembangan sistem informasi berbasis website: Systematic literature review," *Techno.Com*, vol. 21, no. 1, pp. 139–148, 2022.

- [16] I. Sommerville, *Software Engineering*, edisi ke-10. Boston: Pearson, 2016.
- [17] A. S. Rosa dan M. Shalahuddin, *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika, 2021.
- [18] T. Connolly dan C. Begg, *Database Systems: A Practical Approach to Design, Implementation, and Management*, edisi ke-6. Boston: Pearson, 2015.
- [19] R. Abdulloh, *Pemrograman Web untuk Pemula*. Jakarta: Elex Media Komputindo, 2021.
- [20] A. Kadir, *Pengenalan Sistem Informasi*, edisi revisi. Yogyakarta: Andi, 2021.
- [21] T. Sutabri, *Konsep Sistem Informasi*. Yogyakarta: Andi, 2022.
- [22] Uminingsih, M. N. Ichsanudin, M. Yusuf, dan S. Suraya, "Pengujian fungsional perangkat lunak sistem informasi perpustakaan dengan metode black box testing bagi pemula," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 1, no. 2, pp. 1–8, 2022, doi: 10.55123/storage.v1i2.270.