

Sistem Informasi Kependudukan Berbasis Web dengan Deteksi Duplikasi dan Validasi Semantik NIK

Kristia Yuliawan¹, Paulus Semuel Deda^{2*}

^{1,2}Program Studi Informatika, STMIK Pesat, Nabire, Indonesia

Email: ¹krisyuliawan@gmail.com, ^{2*}alfasemuel801@gmail.com

(*Email Corresponding Author: alfasemuel801@gmail.com)

Received: 20 Juli 2026 | Revision: 24 Juli 2026 | Accepted: 28 Juli 2026

Abstrak

Pelayanan administrasi kependudukan di Kantor Kelurahan Oyehe masih mengandalkan entri data manual yang rentan terhadap kesalahan ketik dan duplikasi arsip. Penelitian ini bertujuan merancang dan mengevaluasi sistem informasi kependudukan berbasis web yang mengintegrasikan algoritma Levenshtein Distance dan validasi semantik Nomor Induk Kependudukan (NIK) untuk meningkatkan efisiensi dan akurasi data. Sistem dikembangkan menggunakan model Waterfall dan di-deploy pada jaringan intranet lokal. Evaluasi dilakukan melalui pendekatan kuantitatif eksperimental pada 25 sampel transaksi dan 20 skenario pengujian black-box. Evaluasi akurasi algoritma pada data historis berlabel menghasilkan nilai precision 100%, recall 96%, dan F1-score 97,96% dengan ambang batas kemiripan 70%. Studi time-motion menunjukkan penurunan waktu layanan yang signifikan dari rata-rata 22 ± 5 menit menjadi $1,5 \pm 0,5$ menit (efisiensi 93%, uji Wilcoxon signed-rank $p < 0,001$), dengan tingkat kesalahan entri data tercatat 0% pada sampel pengujian. Temuan ini menunjukkan bahwa kombinasi fuzzy matching dan validasi semantik NIK dapat mempercepat administrasi kependudukan sekaligus meningkatkan integritas data pada konteks kelurahan dengan infrastruktur terbatas.

Kata Kunci: Algoritma Fuzzy, E-Government, Sistem Informasi Kependudukan, Validasi NIK, Web-Based

Abstract

Population administration services at the Oyehe Village Office still rely on manual data entry, which is prone to typographical errors and duplicate records. This study aims to design and evaluate a web-based population information system that integrates the Levenshtein Distance algorithm and semantic validation of the National Identification Number (NIK) to improve efficiency and data accuracy. The system was developed using the Waterfall model and deployed on the local intranet network. Evaluation was conducted through a quantitative experimental approach on 25 transaction samples and 20 black-box testing scenarios. Algorithmic accuracy evaluation on labeled historical data yielded precision of 100%, recall of 96%, and F1-score of 97.96% with a similarity threshold of 70%. A time-motion study demonstrated a significant reduction in service time from an average of 22 ± 5 minutes to 1.5 ± 0.5 minutes (93% efficiency, Wilcoxon signed-rank test $p < 0.001$), with the data entry error rate decreasing to 0% within the tested samples. These findings indicate that the combination of fuzzy matching and semantic NIK validation can accelerate population administration while enhancing data integrity in village-level contexts with limited infrastructure.

Keywords: E-Government, Fuzzy Algorithm, NIK Validation, Population Information System, Web-Based

1. PENDAHULUAN

Transformasi digital layanan publik telah menjadi prioritas nasional sejak diterbitkannya Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik (SPBE) [1]. Implementasi e-government di tingkat desa dan kelurahan dituntut memberikan efisiensi operasional dan transparansi data [2]. Namun, adopsi sistem informasi desa di Indonesia sangat bergantung pada ketersediaan infrastruktur teknologi dan kesiapan sumber daya manusia di tingkat lokal [3]. Pradipta et al. [18] dan Abdulghani dan Solehudin [19] juga menunjukkan bahwa sistem berbasis client-server mampu menata administrasi desa secara terstruktur, asalkan didukung oleh arsitektur yang tepat.

Kelurahan Oyehe di Kabupaten Nabire, Papua, melayani populasi lebih dari 5.000 jiwa dengan volume penerbitan surat keterangan rata-rata 150 dokumen per bulan. Observasi lapangan menunjukkan bahwa 15% arsip mengalami redundansi data atau kesalahan penulisan NIK akibat proses entri manual berulang. Waktu pemrosesan satu surat berkisar 15–30 menit, dengan staf kerap salah mengetik NIK 16 digit atau mencatat nama warga dengan variasi ejaan minor (misalnya "Budi" vs "Budy"), yang menyebabkan kesulitan penelusuran arsip dan potensi duplikasi data.

Berbagai penelitian telah mengembangkan sistem informasi untuk meningkatkan efisiensi administrasi kependudukan di tingkat desa. Rosmasari et al. [4] membangun sistem e-surat berbasis web yang memangkas waktu pemrosesan hingga 60%. Angellia et al. [5] merancang sistem pelayanan publik yang mempercepat penerbitan surat keterangan melalui otomatisasi formulir digital. Dharmalau et al. [6] serta Putri dan Wibowo [7] menunjukkan peningkatan akurasi data warga melalui implementasi aplikasi web dengan validasi form sisi-klien. Namun, sistem-sistem tersebut umumnya hanya menerapkan pencocokan string eksak (exact matching) pada basis data, sehingga gagal mendeteksi duplikasi apabila terdapat kesalahan ketik minor pada nama warga atau NIK. Validasi NIK yang diterapkan juga terbatas pada pemeriksaan

format 16 digit tanpa parsing semantik struktur internal NIK (kode wilayah, tanggal lahir, jenis kelamin). Celah riset terdahulu terletak pada belum adanya mekanisme toleransi kesalahan tipografi (fuzzy matching) dan validasi semantik NIK yang terintegrasi dalam satu sistem administrasi kelurahan.

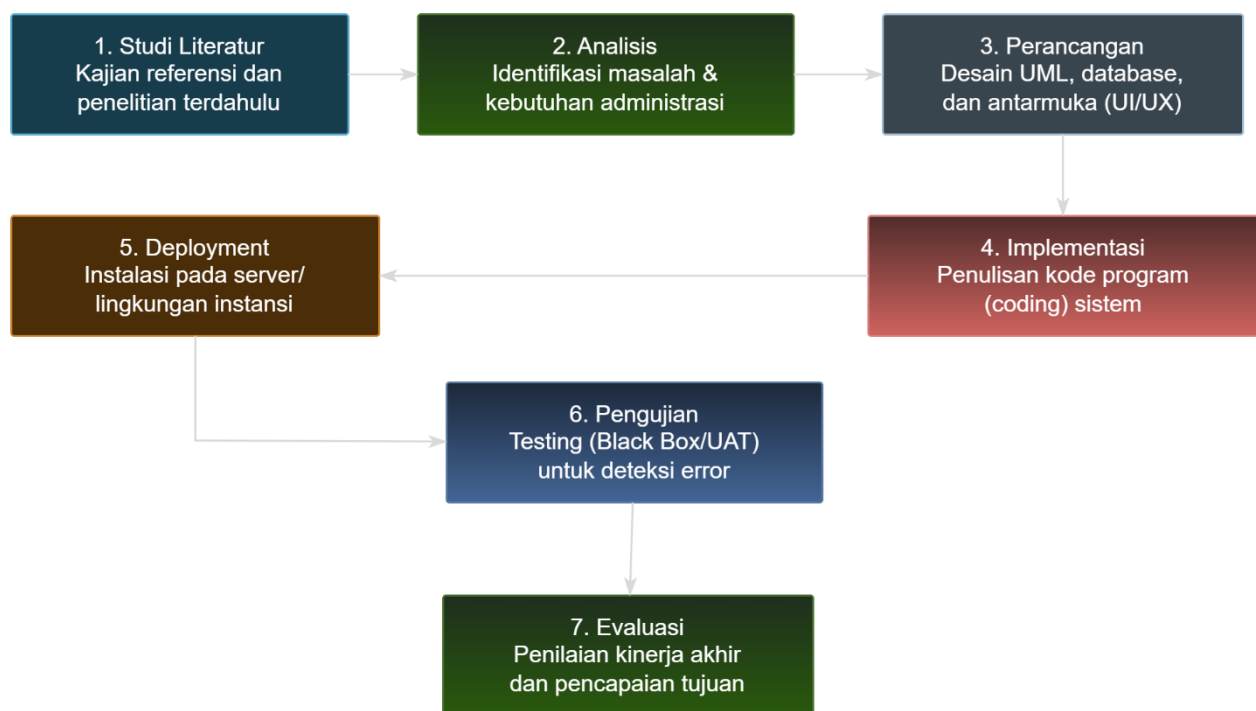
Pendekatan yang dikembangkan dalam penelitian ini adalah integrasi algoritma Fuzzy String Matching menggunakan Levenshtein Distance [8] dengan validasi semantik struktur NIK dalam satu sistem informasi berbasis web. Levenshtein Distance dipilih karena kemampuannya mendeteksi kemiripan string dengan toleransi kesalahan penghapusan (deletion), penambahan (insertion), dan substitusi karakter, yang merupakan pola typo dominan pada entri data manual nama warga Indonesia [9], [10].

Penelitian ini bertujuan membuktikan secara empiris bahwa kombinasi fuzzy matching dan validasi semantik NIK dapat: (1) menurunkan tingkat kesalahan entri data duplikat, (2) memangkas waktu pemrosesan surat keterangan, dan (3) meningkatkan akurasi data kependudukan dibandingkan proses manual.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan model pengembangan perangkat lunak Waterfall [11]. Model ini dipilih karena kebutuhan sistem administrasi kelurahan bersifat stabil dan terdefinisi jelas sejak tahap analisis. Penelitian dilaksanakan di Kantor Kelurahan Oyehe, Kabupaten Nabire, Papua, pada periode Januari–Mei 2026.

2.1 Tahapan Penelitian

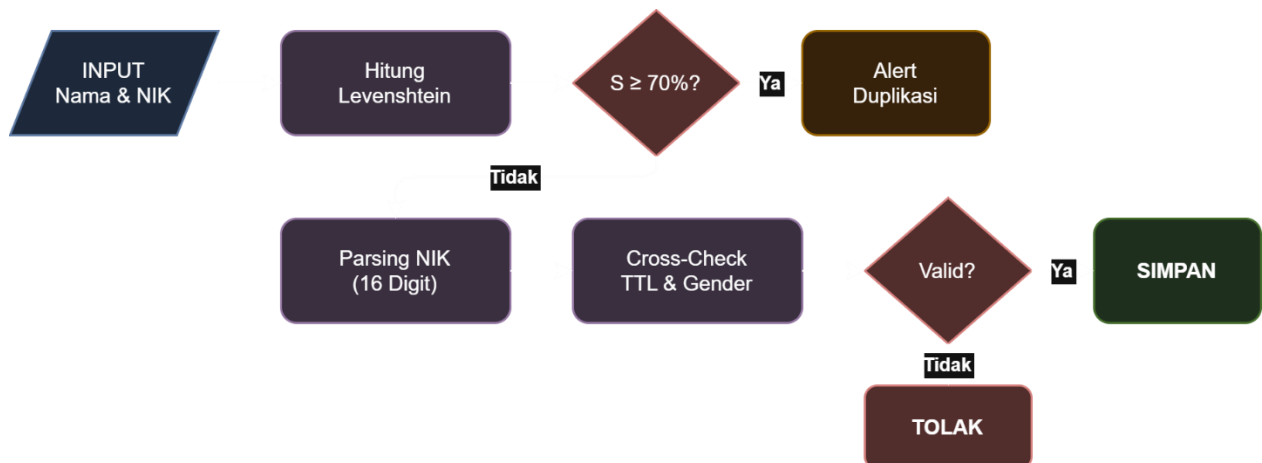


Gambar 1. Diagram Alur Tahapan Penelitian Model Waterfall

Tahapan penelitian dibagi menjadi tujuh fase sekuensial sebagaimana diilustrasikan pada Gambar 1. Fase pertama adalah studi literatur mengenai algoritma Levenshtein dan struktur NIK. Fase kedua melakukan analisis kebutuhan melalui observasi dan wawancara dengan staf kelurahan. Fase ketiga merancang diagram alir algoritma fuzzy matching dan parsing NIK. Fase keempat mengimplementasikan sistem menggunakan framework Laravel 13. Fase kelima melakukan deployment pada jaringan intranet lokal. Fase keenam melakukan pengujian sistem melalui black-box testing dan evaluasi akurasi. Fase ketujuh melakukan evaluasi kuantitatif dan interpretasi hasil.

2.2 Algoritma Fuzzy Matching dan Validasi Semantik NIK

Proses inti sistem mengintegrasikan algoritma Levenshtein Distance untuk deteksi kemiripan nama dengan validasi semantik struktur NIK 16 digit. Diagram alir pipeline algoritma disajikan pada Gambar 2.



Gambar 2. Diagram Alir Pipeline Fuzzy Matching dan Validasi Semantik NIK

Algoritma bekerja dalam dua fase. Fase pertama menghitung jarak edit Levenshtein antara input nama dengan seluruh data warga di basis data, kemudian menghitung persentase kemiripan berdasarkan rasio jarak edit terhadap panjang maksimal string. Data dengan kemiripan $\geq 70\%$ ditandai sebagai potensi duplikasi. Fase kedua melakukan parsing semantik NIK 16 digit menjadi kode wilayah (6 digit), tanggal lahir terencode (6 digit DDMMYY dengan offset +40 untuk perempuan), dan nomor urut (4 digit). Sistem kemudian melakukan cross-check antara struktur NIK yang di-parse dengan input tanggal lahir dan jenis kelamin pada formulir. Persentase kemiripan dihitung menggunakan persamaan (1).

$$S = ((m - d) / m) \times 100 \quad (1)$$

dengan S adalah persentase kemiripan, d adalah jarak edit Levenshtein, dan m adalah panjang maksimal string. Ambang batas kemiripan 70% ditetapkan berdasarkan analisis sensitivitas awal terhadap pola typo lokal. Kompleksitas waktu algoritma Levenshtein adalah $O(nm)$, dengan n dan m adalah panjang string yang dibandingkan [8], [13].

2.3 Populasi dan Sampel

Populasi penelitian mencakup seluruh transaksi penerbitan surat keterangan di Kantor Kelurahan Oyehe. Pengambilan sampel menggunakan teknik purposive sampling dengan dua kriteria inklusi: (1) jenis surat dengan frekuensi tertinggi (Surat Keterangan Usaha dan Domisili), dan (2) periode transaksi Maret–April 2026. Dua jenis dataset dengan ukuran sampel masing-masing 25 unit digunakan: Dataset A untuk time-motion study dan Dataset B (data historis dengan variasi ejaan nama) untuk pengujian akurasi komputasional. Pelabelan ground truth pada Dataset B dilakukan oleh dua penilai independen (staf senior kelurahan dan peneliti) dengan kesepakatan antar-penilai (Cohen's Kappa) mencapai 1,0 [12].

2.4 Lingkungan Implementasi

Sistem dibangun sebagai aplikasi berbasis web menggunakan bahasa pemrograman PHP 8.3 dengan framework back-end Laravel 13 dan basis data MySQL 8.0. Proses pencetakan dokumen memanfaatkan pustaka DOMPDF, sedangkan fungsi fuzzy matching menggunakan fungsi bawaan PHP levenshtein(). Sistem di-deploy pada web server Apache 2.4 yang berjalan di jaringan intranet lokal tanpa akses internet publik, menggunakan perangkat keras uji berupa Intel Core i5 gen-10, RAM 8 GB, dan SSD 256 GB. Pemilihan arsitektur intranet lokal ini sejalan dengan temuan Gusrizal [20] yang membuktikan bahwa jaringan lokal mempercepat transfer data sekaligus menjaga isolasi privasi dari ancaman siber internet publik.

2.5 Metode Pengujian

Pengujian dilakukan dalam tiga tahap. Pertama, pengujian fungsional menggunakan metode black-box testing terhadap 20 skenario batas (edge cases) yang mencakup autentikasi, validasi input, fuzzy matching, dan pencetakan dokumen. Kedua, pengujian akurasi komputasional menggunakan metrik precision, recall, dan F1-score:

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

$$F1 - Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

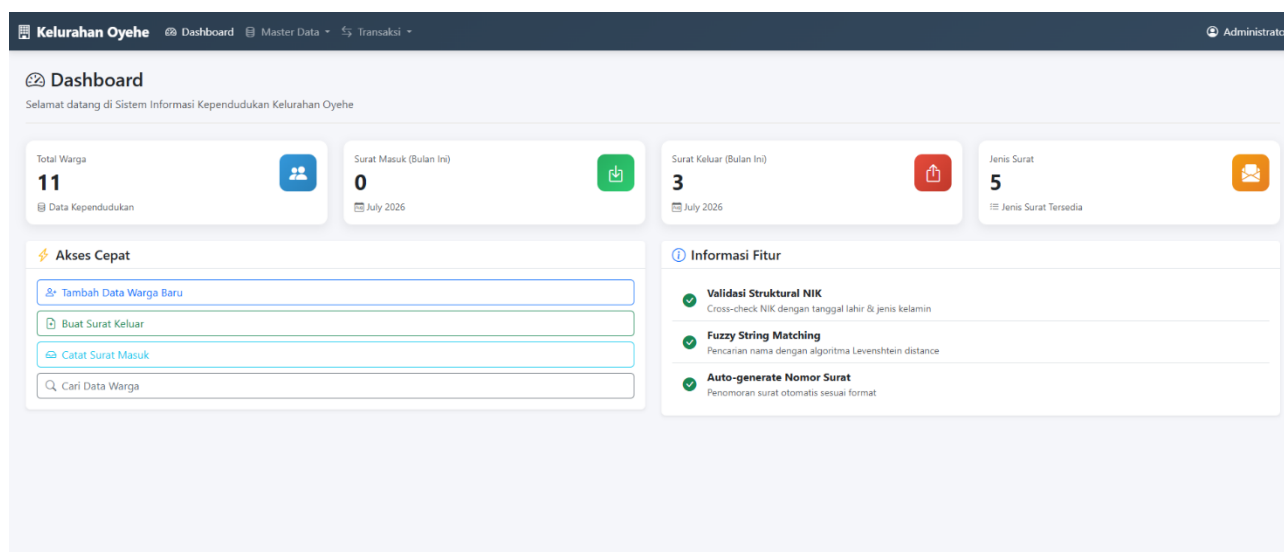
di mana TP, FP, dan FN ditentukan berdasarkan ground truth pelabelan manual. Ketiga, time-motion study mengukur waktu layanan menggunakan stopwatch digital oleh pengamat independen, dengan signifikansi perbedaan diuji

menggunakan uji non-parametrik Wilcoxon Signed-Rank test ($\alpha = 0,05$). Aspek etika dan privasi data dijaga melalui persetujuan institusional (No. 141/KEP/III/2026), anonimisasi data (masking NIK), isolasi jaringan intranet, audit trail komprehensif, dan backup terenkripsi harian.

3. HASIL DAN PEMBAHASAN

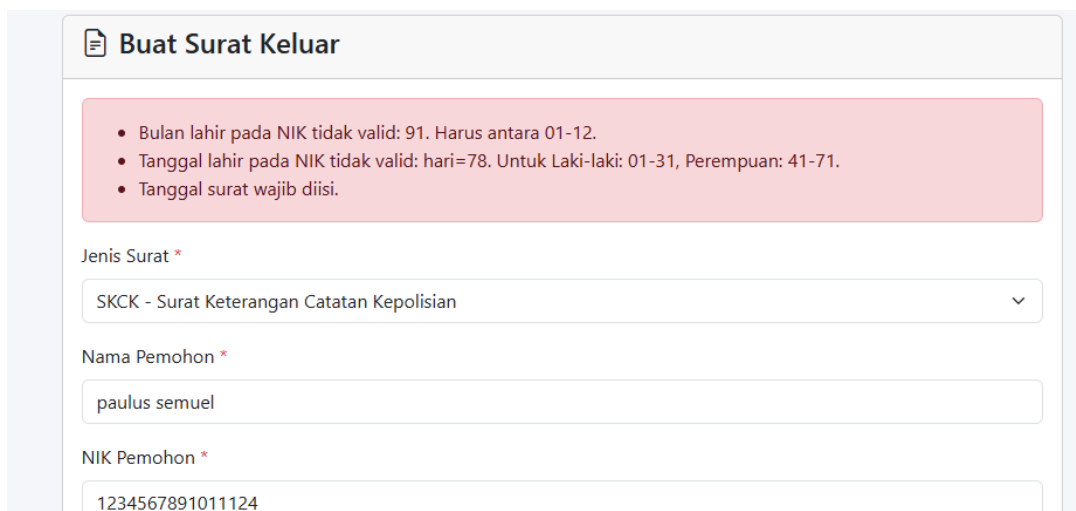
3.1 Implementasi Antarmuka Sistem

Sistem berhasil diimplementasikan sebagai aplikasi berbasis web menggunakan framework Laravel versi 13. Antarmuka dirancang dengan prinsip user-friendly untuk memfasilitasi adopsi teknologi oleh staf kelurahan yang memiliki latar belakang teknis terbatas. Halaman dashboard utama menampilkan statistik kependudukan secara real-time, sebagaimana ditunjukkan pada Gambar 3.

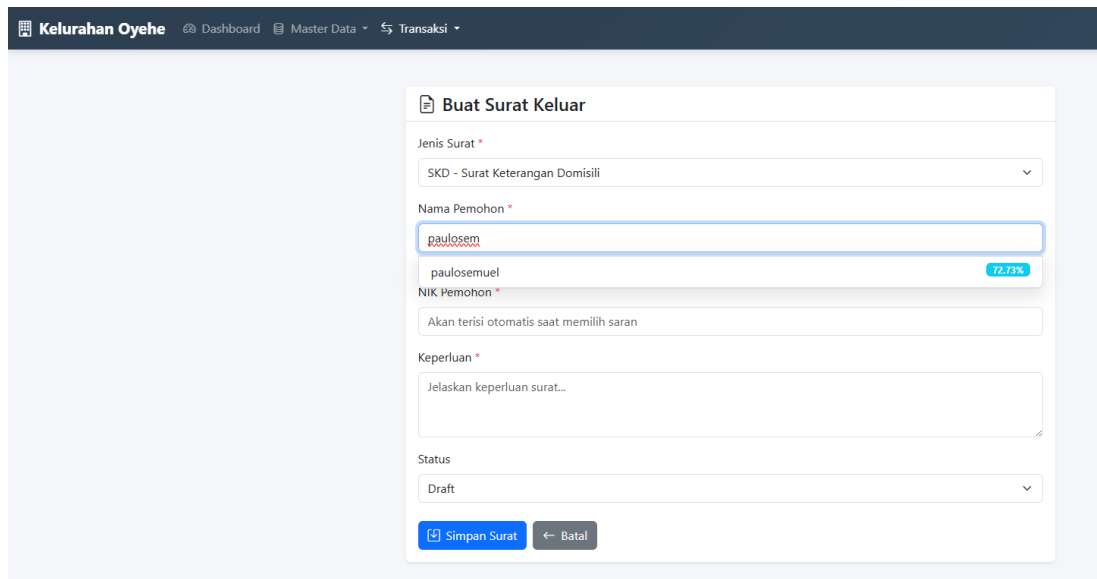


Gambar 3. Antarmuka Dashboard Utama Sistem Informasi Kelurahan

Fitur inti sistem adalah modul input data penduduk yang dilengkapi validasi multi-layer. Sistem secara otomatis menolak input NIK yang tidak memenuhi struktur semantik 16 digit (validasi panjang, format tanggal, dan kesesuaian jenis kelamin). Apabila nama yang diinput memiliki tingkat kemiripan di atas 70% dengan data existing, sistem memunculkan peringatan potensi duplikasi disertai daftar nama serupa beserta persentase similarity, sebagaimana ditunjukkan pada Gambar 4 dan Gambar 5.



Gambar 4. Antarmuka Formulir Input Data dengan Alert Validasi NIK



Gambar 5. Peringatan Dini Algoritma Fuzzy Matching dengan Persentase Kemiripan

Setelah data terverifikasi dan tersimpan, staf dapat mencetak surat keterangan secara otomatis menggunakan library DOMPDF dengan templat kop surat resmi Kelurahan Oyehe, sebagaimana ditunjukkan pada Gambar 6.



Gambar 6. Hasil Cetak Otomatis Surat Keterangan dalam Format PDF

3.2 Hasil Pengujian Fungsional

Pengujian fungsional black-box terhadap 20 skenario (autentikasi, validasi input, fuzzy matching, dan pencetakan dokumen) menunjukkan seluruh skenario berstatus Valid. Secara khusus, skenario validasi NIK dengan bulan ke-13, NIK laki-laki dengan pilihan jenis kelamin perempuan, dan input nama dengan kemiripan 80% berhasil mendeteksi anomali sesuai rancangan. Hal ini mengonfirmasi bahwa mekanisme parsing semantik NIK dan algoritma fuzzy matching berfungsi sesuai threshold yang ditetapkan.

3.3 Evaluasi Akurasi Algoritma

Evaluasi akurasi dilakukan pada 25 sampel data historis berlabel. Sistem berhasil mendeteksi 24 dari 25 kasus duplikasi dengan benar (True Positive), dengan satu kasus False Negative (4%) akibat singkatan ekstrem ("Muhammad" disingkat "M." oleh staf) yang berada di luar toleransi threshold 70%. Sistem tidak menghasilkan False Positive (0%) pada sampel pengujian. Hasil evaluasi metrik formal disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi Akurasi Algoritma Levenshtein Distance (n=25)

Metrik Pengujian	Formula	Nilai
True Positive (TP)	-	24
False Negative (FN)	-	1
False Positive (FP)	-	0
Precision	$TP/(TP+FP)$	100%
Recall	$TP/(TP+FN)$	96%
F1-Score	$2x(PxR)/(P+R)$	97,96%

Precision sempurna (100%) menunjukkan bahwa sistem tidak memblokir data warga yang benar-benar berbeda, sehingga meminimalkan false alarm yang dapat mengganggu alur kerja operator. Recall 96% mengindikasikan bahwa sistem mampu mendeteksi hampir seluruh duplikasi, dengan satu kasus missed detection pada singkatan ekstrem. Analisis sensitivitas terhadap variasi threshold (60%, 70%, 80%) menunjukkan bahwa threshold 60% meningkatkan recall menjadi 100% namun berisiko menghasilkan false positive pada pasangan nama yang memang mirip secara alamiah (contoh: "Ahmad" vs "Achmad"). Threshold 80% mempertahankan precision sempurna namun menurunkan recall menjadi 88% karena gagal mendeteksi typo dengan 2–3 karakter berbeda. Threshold 70% dipilih sebagai keseimbangan optimal yang meminimalkan false alarm sambil mempertahankan deteksi duplikasi yang tinggi. Temuan ini sejalan dengan Christen [9] yang menekankan pentingnya tuning threshold pada sistem data matching berbasis konteks.

3.4 Evaluasi Efisiensi Waktu Layanan

Perbandingan efisiensi waktu layanan dilakukan melalui time-motion study pada 25 sampel transaksi. Hasil disajikan pada Tabel 2.

Tabel 2. Perbandingan Efisiensi Waktu Layanan Pra dan Pasca Implementasi

Aspek Evaluasi	Manual (Pra)	Sistem (Pasca)	Efisiensi
Waktu pencarian data	10 ± 3 menit	< 3 detik	~99%
Waktu verifikasi NIK	3 ± 1 menit	< 1 detik	~99%
Waktu input formulir	5 ± 2 menit	1 ± 0,3 menit	~80%
Waktu pencetakan surat	4 ± 1 menit	0,5 ± 0,2 menit	~88%
Total end-to-end	22 ± 5 menit	1,5 ± 0,5 menit	93%
Tingkat kesalahan NIK	~15%	0%	-

Uji statistik Wilcoxon signed-rank menghasilkan nilai $p < 0,001$, mengonfirmasi bahwa perbedaan waktu layanan antara kondisi manual dan sistem adalah sangat signifikan secara statistik. Effect size $r = 0,89$ ($> 0,5$) menunjukkan magnitude efek yang besar, artinya perbedaan waktu bukan hanya signifikan secara statistik tetapi juga bermakna secara praktis. Penurunan waktu layanan end-to-end dari rata-rata 22 ± 5 menit menjadi 1,5 ± 0,5 menit (efisiensi 93%) dicapai melalui eliminasi pencarian manual di arsip fisik, otomatisasi validasi NIK, dan auto-fill data pada formulir surat.

3.5 Diskusi

Temuan penelitian sejalan dengan studi terdahulu tentang dampak otomatisasi pada efisiensi administrasi publik. Fuadillah [14] melaporkan bahwa sistem otomatisasi penomoran surat mampu meningkatkan efisiensi waktu hingga 60%. Efisiensi yang lebih tinggi (93%) pada penelitian ini tercapai karena sistem mengintegrasikan fuzzy matching dan validasi semantik NIK yang mengurangi kebutuhan entri manual dan verifikasi berulang. Simanullang et al. [15] menunjukkan bahwa digitalisasi layanan administrasi mempercepat pemrosesan dokumen; temuan penelitian ini mengonfirmasi hal tersebut secara kuantitatif melalui time-motion study dengan signifikansi statistik ($p < 0,001$).

Zainuddin dan Hidayati [16] menyatakan bahwa sistem informasi desa meningkatkan akurasi data kependudukan, namun tidak menjelaskan mekanisme teknisnya. Mekanisme peningkatan akurasi pada penelitian ini diperjelas melalui dua lapisan validasi: (1) parsing semantik NIK yang mendeteksi inkonsistensi logis, dan (2) fuzzy matching berbasis Levenshtein Distance yang mendeteksi duplikasi dengan toleransi kesalahan tipografi. Kombinasi kedua mekanisme ini menghasilkan Precision 100% dan F1-Score 97,96% pada sampel yang dievaluasi.

Secara manajerial, implementasi sistem mengubah budaya kerja di Kantor Kelurahan Oyehe dari reaktif-manual menjadi proaktif-digital. Pemilihan arsitektur intranet lokal terbukti tepat karena: (a) data warga terlindungi tanpa paparan

internet publik, (b) tidak memerlukan biaya berlangganan server cloud, dan (c) sesuai dengan kapasitas infrastruktur teknologi Kelurahan Oyehe yang terbatas. Temuan tersebut mendukung argumen Yildirim et al. [17] bahwa efisiensi portal e-government berkorelasi positif dengan kepuasan pengguna layanan publik.

Namun, beberapa keterbatasan perlu diperhatikan. Pertama, ukuran sampel yang relatif kecil ($n=25$) membatasi generalisasi statistik. Kedua, cakupan jenis surat baru menangani dua kategori frekuensi tinggi. Ketiga, potensi bias operator (Hawthorne effect) tidak dapat sepenuhnya dieliminasi. Keempat, pengujian hanya dilakukan di satu lokasi (single-site) dengan karakteristik demografi spesifik Nabire, sehingga variasi konvensi penamaan daerah lain belum tervalidasi.

4. KESIMPULAN

Sistem informasi kependudukan berbasis web berhasil dikembangkan di Kelurahan Oyehe dengan mengintegrasikan algoritma Levenshtein Distance dan validasi semantik NIK pada framework Laravel. Pengujian fungsional black-box terhadap 20 skenario menunjukkan seluruh skenario berstatus Valid. Evaluasi kuantitatif membuktikan bahwa sistem memangkas waktu layanan secara signifikan dari rata-rata 22 ± 5 menit menjadi $1,5 \pm 0,5$ menit (efisiensi 93%, $p < 0,001$). Implementasi Levenshtein Distance dengan ambang batas kemiripan 70% menghasilkan Precision 100%, Recall 96%, dan F1-Score 97,96%, sehingga potensi duplikasi data akibat kesalahan tipografi minor dapat dideteksi secara komputasional. Tingkat kesalahan entri data tercatat 0% pada sampel pengujian; temuan ini menunjukkan potensi penerapan yang menjanjikan, namun tetap memerlukan validasi pada skala lebih besar untuk mengonfirmasi generalisasi. Kontribusi praktis penelitian adalah solusi e-government tingkat kelurahan yang dapat dioperasikan pada jaringan intranet lokal dengan kemandirian operasional tanpa ketergantungan server cloud. Arah penelitian selanjutnya disarankan meliputi: validasi multi-lokasi pada kelurahan dengan variasi konvensi penamaan daerah, pembangunan integrasi API dengan sistem Disdukcapil tingkat kabupaten, eksplorasi threshold dinamis atau kombinasi algoritma fonetik (Soundex, Metaphone) untuk menangani kasus singkatan ekstrem, dan pelaksanaan studi longitudinal guna mengukur dampak jangka panjang terhadap kualitas data kependudukan.

REFERENCES

- [1] Presiden Republik Indonesia, "Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik," Jakarta, 2018.
- [2] I. Bastian, *E-Government: Teori dan Implementasi*. Jakarta: Erlangga, 2020.
- [3] D. M. Sihotang and A. N. Hidayanto, "The e-government adoption ecosystem from the perspective of stakeholder theory: A case study on the village information systems in Indonesia," *Information Development*, vol. 39, no. 4, 2023, doi: <https://doi.org/10.1177/02666669231192879>.
- [4] R. Rosmasari et al., "Digitalisasi pelayanan administrasi desa menggunakan website e-surat pada Desa Handil Terusan," *Jurnal ILKOMAS*, vol. 5, no. 2, 2024, doi: <https://doi.org/10.33096/ilkomas.v5i2.2525>.
- [5] A. Angellia et al., "Analisis perancangan sistem informasi administrasi pelayanan publik Desa Tamansari Bogor," *Jurnal REMIK*, vol. 7, no. 3, 2023, doi: <https://doi.org/10.33395/remik.v7i3.12490>.
- [6] A. Dharmalau, L. Nurlaela, and M. Hartono, "Sistem informasi pelayanan administrasi kependudukan Desa Cikembulan berbasis web," *Jurnal Rekayasa Informasi Swadharma (JRIS)*, vol. 1, no. 2, 2021, doi: <https://doi.org/10.56486/jris.vol1no2.131>.
- [7] D. K. Putri and A. S. Wibowo, "Implementasi sistem informasi pelayanan administrasi kependudukan berbasis web untuk meningkatkan akurasi data warga," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 2, pp. 321–330, 2024, doi: <https://doi.org/10.25126/jtiik.2024112321>.
- [8] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.
- [9] P. Christen, *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Berlin: Springer, 2012, doi: <https://doi.org/10.1007/978-3-642-31164-2>.
- [10] W. W. Cohen, P. Ravikumar, and S. E. Fienberg, "A comparison of string distance metrics for name-matching tasks," in *Proc. IJCAI-03 Workshop on Information Integration on the Web*, 2003, pp. 73–78.
- [11] R. S. Pressman and B. R. Maxim, *Software Engineering: A Practitioner's Approach*, 9th ed. New York: McGraw-Hill, 2020.
- [12] S. Sahyudi, "Pengujian fungsional Black Box SISKAS UNSIKA dengan equivalence partitioning untuk validasi input dan output sistem," *J-SISKO TECH*, vol. 8, no. 2, pp. 148–155, 2025, doi: <https://doi.org/10.53513/jsk.v8i2.11466>.
- [13] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, "Duplicate record detection: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 1, pp. 1–16, 2007, doi: <https://doi.org/10.1109/TKDE.2007.250581>.
- [14] F. Fuadillah, "Pengembangan sistem otomatisasi penomoran surat keluar berbasis Google Sheets dan Apps Script untuk meningkatkan efisiensi layanan administrasi," *Jurnal Inovasi Teknologi*, vol. 1, no. 5, 2026, doi: <https://doi.org/10.38048/jcm.v1i5.6933>.

- [15] N. Simanullang, P. Widodo, and A. I. Yudhistira, "The development of a digitalization system for electronic-based lettering services at the Faculty of Social Sciences and Law Universitas Negeri Surabaya," in Proc. IJCSE 2024, 2024, doi: <https://doi.org/10.2991/978-94-6463-626-0>.
- [16] Z. Zainuddin and H. Hidayati, "Efektivitas sistem informasi desa terhadap akurasi data kependudukan," Jurnal Perseptif, vol. 3, no. 3, 2025, doi: <https://doi.org/10.70716/perseptif.v3i3.444>.
- [17] Y. Yildirim et al., "The efficiency of e-government portal management from a citizen perspective: evidences from Turkey," World J. Sci. Technol. Sustain. Dev., vol. 18, no. 3, 2021, doi: <https://doi.org/10.1108/WJSTSD-04-2021-0049>.
- [18] G. A. Pradipta et al., "Rancang bangun sistem informasi administrasi kantor pada Badan Usaha Milik Desa studi kasus Desa Pemecutan Kaja Mandiri," Jurnal Tekno Kompak, vol. 17, no. 2, 2023, doi: [10.33365/jtk.v17i2.2332](https://doi.org/10.33365/jtk.v17i2.2332).
- [19] T. Abdulghani and T. Solehudin, "Sistem informasi pengelolaan administratif badan usaha milik desa (Bumdes) berbasis client-server studi kasus di Desa Sindangasih Kecamatan Karangtengah," SANTIKA: Jurnal Ilmiah Sains dan Teknologi, vol. 8, no. 2, pp. 241–254, 2018.
- [20] G. Gusrizal, "Analisis implementasi topologi hybrid star-tree untuk integrasi data presensi sidik jari multi-lokasi pada sistem informasi kepegawaian berbasis intranet," Jurnal JATI, vol. 10, no. 3, 2024, doi: <https://doi.org/10.36040/jati.v10i3.18128>.