JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

Perancangan dan Implementasi Sistem PPDB Online Berbasis Web dengan Validasi Dokumen Menggunakan OCR dan Machine Learning di SMAN 3 Kilo Kabupaten Dompu

M. Rajuliadin[1], Cendra Harmom[2]

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: [1]mrajuliadin@gmail.com, [2]dosen02677@unpam.ac.id

Abstrak–Proses Penerimaan Peserta Didik Baru (PPDB) di SMAN 3 Kilo, Kabupaten Dompu, masih menggunakan pemeriksaan dokumen secara manual sehingga berpotensi meningkatkan beban kerja panitia dan menimbulkan kesalahan verifikasi. Penelitian ini bertujuan merancang dan mengimplementasikan sistem PPDB berbasis web dengan validasi dokumen menggunakan Optical Character Recognition (OCR), validasi berbasis aturan, dan algoritma Isolation Forest. Penelitian menggunakan pendekatan Research and Development (R&D) dengan metode pengembangan Rapid Application Development (RAD). Dokumen yang diperiksa meliputi Kartu Keluarga (KK) dan Ijazah atau Surat Keterangan Lulus (SKL). Karena dokumen asli yang tersedia hanya sekitar 10 berkas, pengembangan model menggunakan 350 sampel yang terdiri atas 300 data normal dan 50 data anomali. Sistem menghasilkan status auto_lolos atau perlu_review sebagai indikator pemeriksaan dokumen oleh panitia. Hasil pengujian pipeline menunjukkan bahwa sistem dapat membedakan dokumen berdasarkan kondisi dan kesesuaian data. Evaluasi Isolation Forest menggunakan 10 data uji menghasilkan nilai precision, recall, dan F1-score masing-masing sebesar 80%. Hasil tersebut masih terbatas pada data pengujian dan belum dapat digeneralisasikan terhadap dokumen PPDB secara luas. Keputusan akhir penerimaan tetap menjadi kewenangan panitia.

Kata Kunci: PPDB Online, Optical Character Recognition, Validasi Dokumen, Isolation Forest, Machine Learning.

Abstract–*The New Student Admission (PPDB) process at SMAN 3 Kilo, Dompu Regency, still relies on manual document verification, which may increase the admission committee's workload and the risk of verification errors. This study aims to design and implement a web-based PPDB system with document validation using Optical Character Recognition (OCR), rule-based validation, and the Isolation Forest algorithm. The research employs a Research and Development (R&D) approach with Rapid Application Development (RAD) as the system development method. The documents examined are Family Cards (Kartu Keluarga/KK) and Diplomas or Graduation Certificates (Ijazah/Surat Keterangan Lulus/SKL). Due to the limited availability of original documents, approximately 10 files, the model was developed using 350 samples consisting of 300 normal data samples and 50 anomaly samples. The system generates auto_lolos or perlu_review statuses as indicators to support document verification by the admission committee. Pipeline testing shows that the system can distinguish documents based on their condition and data consistency. Evaluation of the Isolation Forest model using 10 test samples resulted in precision, recall, and F1-score values of 80% each. These results are limited to the test dataset and cannot yet be generalized to the broader range of PPDB documents. Final admission decisions remain under the authority of the admission committee.*

Keywords: *Online Student Admission, Optical Character Recognition, Document Validation, Isolation Forest, Machine Learning.*

1. PENDAHULUAN

Penerimaan Peserta Didik Baru (PPDB) merupakan proses administratif yang meliputi pendataan, pemeriksaan persyaratan, dan penetapan hasil penerimaan. Pengelolaan PPDB berbasis web telah dibahas oleh Budiarto et al. (2022) dan Yudahana et al. (2023), sedangkan Kiayi et al. (2025) mengkaji penerapan metode Rapid Application Development (RAD) dalam aplikasi pendaftaran siswa baru. Di SMAN 3 Kilo, Kabupaten Dompu, proses verifikasi masih dilakukan secara manual. Sekolah yang berada di wilayah kategori 3T (Terdepan, Terluar, dan Tertinggal) memiliki keterbatasan infrastruktur teknologi dan sumber daya manusia, dengan sekitar 50–60 pendaftar setiap tahun. Panitia memeriksa kelengkapan dokumen Kartu Keluarga (KK) dan Ijazah atau Surat Keterangan Lulus (SKL), termasuk kesesuaian Nomor Induk Kependudukan (NIK) dan identitas pendaftar. Pemeriksaan berpotensi membutuhkan waktu dan tenaga lebih besar ketika dokumen buram atau memuat informasi tidak konsisten.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

Untuk membantu pemeriksaan, penelitian ini mengintegrasikan Optical Character Recognition (OCR), validasi berbasis aturan (rule-based validation), dan algoritma Isolation Forest. OCR digunakan untuk mengekstraksi teks dokumen, sebagaimana dibahas oleh Toha dan Triayudi (2022) serta Darpito et al. (2025). Hasil ekstraksi diperiksa berdasarkan keberadaan teks, format NIK, dan kesesuaian informasi; pendekatan validasi dokumen berbasis AI juga dibahas oleh Fedhira dan Prianto (2025). Isolation Forest digunakan untuk mengidentifikasi pola dokumen menyimpang, dengan penerapan deteksi anomali yang dikaji oleh Waspada et al. (2020), Mutmainah dan Yustanti (2024), serta Ohyver et al. (2025). Ketiga pendekatan tersebut mendukung pemeriksaan awal dan tidak menjadi penentu tunggal keabsahan dokumen.

Pengembangan model menghadapi keterbatasan karena hanya tersedia sekitar sepuluh dokumen asli dan sekolah belum memiliki sistem PPDB digital sebelumnya. Oleh karena itu, pelatihan model didukung augmentasi dan data sintetis berdasarkan karakteristik dokumen yang tersedia. Strategi ini memungkinkan pengembangan dan pengujian awal, tetapi belum mewakili seluruh variasi dokumen PPDB nyata. Penelitian menggunakan pendekatan Research and Development (R&D) dengan metode RAD yang mencakup perencanaan kebutuhan, perancangan, dan implementasi. Perancangan perangkat lunak mengacu pada Rosa dan Shalahuddin (2024), sedangkan penerapan RAD pada sistem pendaftaran peserta didik dibahas oleh Yudahana et al. (2023) dan Kiayi et al. (2025). Sistem dikembangkan menggunakan Laravel untuk aplikasi web dan Python untuk pemrosesan dokumen.

Penelitian ini bertujuan merancang dan mengimplementasikan sistem PPDB berbasis web yang mengintegrasikan OCR, validasi berbasis aturan, dan Isolation Forest untuk memeriksa dokumen KK dan Ijazah/SKL. Sistem menghasilkan indikator auto_lolos atau perlu_review sebagai rekomendasi bagi panitia. Cakupan penelitian meliputi pendaftaran, pengunggahan dokumen, ekstraksi teks, dan pemeriksaan awal, tanpa mencakup penilaian akademik, pemeringkatan, atau penentuan kelulusan otomatis. Keputusan akhir tetap menjadi kewenangan panitia PPDB, sedangkan evaluasi Isolation Forest merupakan studi kelayakan karena keterbatasan dokumen asli dan penggunaan data sintetis.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan Research and Development (R&D) dengan metode Rapid Application Development (RAD), yang meliputi requirements planning, RAD design workshop, dan implementation. Penerapan RAD pada sistem pendaftaran berbasis web dibahas oleh Yudahana et al. (2023) dan Kiayi et al. (2025).

Penelitian dilaksanakan di SMAN 3 Kilo, Kabupaten Dompu, Nusa Tenggara Barat, dengan objek dokumen Kartu Keluarga (KK) dan Ijazah atau Surat Keterangan Lulus (SKL). Data dikumpulkan melalui observasi, wawancara, dokumentasi, dan studi pustaka. Sampel dipilih secara purposive sampling berdasarkan kondisi dokumen, yaitu jelas, buram, dan terpotong. Data historis berjumlah sekitar 10 dokumen asli, kemudian dilengkapi augmentasi dan data sintetis berdasarkan parameter statistik dokumen tersebut. Dataset pelatihan Isolation Forest terdiri atas 350 sampel, yaitu 300 normal dan 50 anomali. Penggunaan data sintetis ditujukan untuk mengatasi keterbatasan data (cold-start problem) dan mendukung pengujian awal, bukan menggantikan pengujian dengan dokumen nyata dalam jumlah memadai.

Pengembangan sistem melalui tiga tahap RAD. Requirements planning mengidentifikasi kebutuhan pendaftaran daring, pengisian data, pengunggahan dokumen, validasi otomatis, dan pemantauan status. RAD design workshop menghasilkan rancangan arsitektur dan basis data yang mengintegrasikan aplikasi web Laravel dan PHP dengan layanan Python melalui basis data bersama. Perancangan mengacu pada prinsip rekayasa perangkat lunak Rosa dan Shalahuddin (2024).

Tahap implementation menerapkan tiga lapisan validasi. Lapisan pertama menggunakan OCR EasyOCR dengan model pretrained tanpa pelatihan ulang. PDF dikonversi menjadi citra menggunakan PyMuPDF, sedangkan gambar diproses dengan PIL. Pemrosesan PDF dibatasi pada halaman pertama. Hasil OCR berupa teks dan nilai confidence digunakan untuk mengukur keyakinan pengenalan karakter. Penerapan OCR pada dokumen digital dibahas oleh Toha dan Triayudi (2022), sedangkan perbandingan pustaka OCR dikaji oleh Darpito et al. (2025).

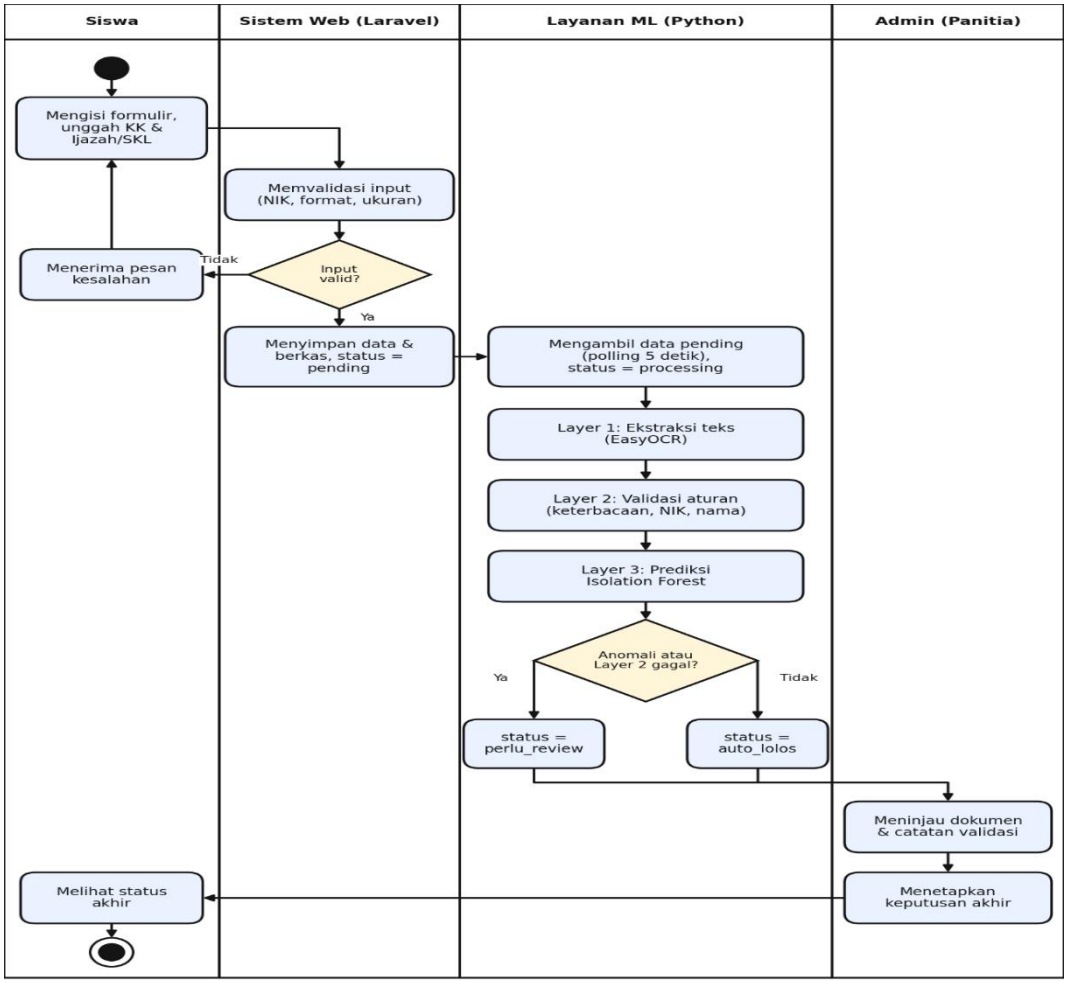
Lapisan kedua menerapkan rule-based validation terhadap keberadaan teks, keterbacaan OCR, kecocokan NIK, dan kesesuaian kata pertama nama pendaftar dengan KK. Keterbacaan

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

terpenuhi apabila teks berhasil diekstraksi dengan rata-rata confidence minimal 0,2. Pencocokan NIK menggunakan normalisasi huruf "o" menjadi angka nol serta "i" atau "l" menjadi angka satu. Ketidaksesuaian nama dicatat sebagai peringatan, bukan langsung menggagalkan validasi. Pemeriksaan deterministik ini memudahkan penelusuran alasan ketidaksesuaian oleh panitia (Fedhira dan Prianto, 2025).

Lapisan ketiga menggunakan Isolation Forest untuk mendeteksi anomali berdasarkan empat fitur: rata-rata confidence OCR, confidence terendah, panjang teks hasil ekstraksi, dan jumlah peringatan validasi aturan. Metode unsupervised learning ini mendeteksi anomali melalui pemisahan data secara acak (Waspada et al., 2020; Mutmainah dan Yustanti, 2024). Model dilatih secara offline menggunakan 350 sampel sintetis dan disimpan sebagai model.pkl, dengan konfigurasi n_estimators=100, max_samples=auto, contamination=0.15, dan random_state=42. Karena keterbatasan dokumen asli, penerapan model merupakan studi kelayakan (feasibility study) dan belum mewakili kinerja pada seluruh variasi dokumen nyata.

Integrasi aplikasi web dan layanan Python dilakukan melalui basis data bersama. Setelah formulir dan dokumen dikirim, aplikasi menyimpan pendaftaran dengan status pending. Worker Python memeriksa basis data setiap lima detik, mengubah status menjadi processing, kemudian menjalankan OCR, validasi aturan, dan prediksi Isolation Forest. Alur hubungan siswa, aplikasi web, layanan pemrosesan, dan panitia PPDB ditunjukkan melalui Activity Diagram pada Gambar 1.



Gambar 1. Activity Diagram Pipeline Validasi Dokumen

Gambar 1 memperlihatkan alur pengisian formulir, pengunggahan dokumen, pemeriksaan otomatis, hingga peninjauan admin. Sistem menghasilkan status auto_lolos atau perlu_review

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

sebagai rekomendasi, sedangkan keputusan akhir penerimaan tetap menjadi kewenangan panitia PPDB.

Pengujian fungsional menggunakan metode Black Box Testing untuk memeriksa registrasi, formulir, pengunggahan dokumen, validasi, tampilan informasi pendaftar, dan pengelolaan keputusan admin. Pengujian pipeline mencakup variasi kondisi serta kesesuaian dokumen. Kinerja Isolation Forest dievaluasi menggunakan confusion matrix dengan indikator precision, recall, dan F1-score. Hasilnya merupakan evaluasi awal karena keterbatasan dokumen asli dan penggunaan data sintetis, sehingga belum dapat digeneralisasi pada seluruh dokumen PPDB nyata.

3. ANALISA DAN PEMBAHASAN

Perancangan sistem Penerimaan Peserta Didik Baru (PPDB) berbasis web dengan mekanisme validasi dokumen menggunakan Optical Character Recognition (OCR), validasi berbasis aturan (rule-based validation), dan algoritma Isolation Forest menghasilkan sejumlah komponen yang saling terintegrasi. Analisis dan pembahasan difokuskan pada lingkungan pengembangan, implementasi aplikasi web, penerapan layanan machine learning, pengelolaan basis data, serta antarmuka sistem. Hasil pengujian fungsional dan evaluasi model juga dianalisis untuk mengetahui kesesuaian sistem terhadap kebutuhan penelitian serta mengidentifikasi keterbatasan penerapannya.

3.1 Implementasi Sistem

Implementasi sistem merupakan penerapan rancangan menjadi aplikasi PPDB berbasis web untuk mengelola pendaftaran dan pemeriksaan dokumen di SMAN 3 Kilo, Kabupaten Dompu. Sistem mengintegrasikan aplikasi web berbasis Laravel dengan layanan pemrosesan dokumen menggunakan Python melalui basis data bersama. Implementasi mencakup lingkungan pengembangan, aplikasi web, layanan machine learning, basis data, dan antarmuka pengguna.

3.1.1 Lingkungan Pengembangan

Pengembangan dan pengujian sistem dilakukan menggunakan laptop Dell Latitude 7390 sebagai perangkat utama untuk menjalankan aplikasi web, basis data, dan layanan pemrosesan Python. Spesifikasi perangkat keras yang digunakan disajikan pada Tabel 1.

Tabel 1. Spesifikasi Perangkat Keras

Komponen	Spesifikasi
Perangkat	Laptop Dell Latitude 7390
Prosesor	Intel Core i5-7300U @ 2,60 GHz (2,71 GHz)
Memori (RAM)	8 GB (7,88 GB dapat digunakan)
Kartu grafis	Intel HD Graphics 620 (128 MB)
Penyimpanan	SSD Samsung MZVLB256HAHQ, kapasitas 238 GB
Sistem operasi	Windows 64-bit (prosesor berbasis x64)

Perangkat menggunakan prosesor Intel Core i5-7300U, RAM 8 GB, dan kartu grafis terintegrasi Intel HD Graphics 620 tanpa GPU khusus. Oleh karena itu, pemrosesan EasyOCR dijalankan tanpa akselerasi GPU.

Sistem menggunakan Laravel dan PHP untuk backend, Laravel Blade serta Tailwind CSS untuk antarmuka, dan Python untuk pemrosesan dokumen. Teknologi pendukung lainnya meliputi SQLite dan MySQL, Laravel Breeze, EasyOCR, PyMuPDF, scikit-learn Isolation Forest, serta joblib. Rincian teknologi dan fungsinya disajikan pada Tabel 2.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

Tabel 2. Teknologi yang Digunakan dalam Implementasi Sistem

Komponen	Teknologi	Keterangan
Backend web	Laravel (PHP) versi LTS	Menggunakan Blade, bukan single page application
Frontend	Laravel Blade dan Tailwind CSS	Antarmuka dirender di sisi server
Basis data	SQLite dan MySQL	SQLite untuk pengembangan dan MySQL untuk produksi; digunakan bersama oleh Laravel dan Python
Autentikasi	Laravel Breeze	Mendukung dua peran, yaitu siswa dan admin
Mesin OCR	EasyOCR	Mendukung Bahasa Indonesia dan Inggris, dijalankan tanpa GPU
Dukungan PDF	PyMuPDF	Mengonversi halaman pertama PDF menjadi citra untuk diproses OCR
Model machine learning	scikit-learn Isolation Forest	Menggunakan parameter n_estimators=100 dan contamination=0.15
Penyimpanan model	joblib	Model disimpan dalam berkas model.pkl
Antrean pemrosesan	Polling basis data oleh Python	Pemeriksaan data setiap lima detik tanpa Redis atau message broker

Aplikasi Laravel dan layanan Python menggunakan basis data bersama untuk bertukar informasi. Layanan Python memeriksa data setiap lima detik melalui mekanisme polling tanpa Redis atau message broker. EasyOCR digunakan untuk ekstraksi teks, sedangkan Isolation Forest menjalankan deteksi anomali berdasarkan fitur hasil ekstraksi dan pemeriksaan aturan. Model disimpan dalam berkas model.pkl agar dapat dimuat saat proses prediksi.

3.1.2 Implementasi Aplikasi Web

Aplikasi web dikembangkan menggunakan Laravel dan Laravel Breeze untuk autentikasi dengan dua peran pengguna, yaitu siswa dan admin, yang memiliki hak akses berbeda. Penggunaan Laravel dalam pengembangan sistem informasi berbasis web dibahas oleh Budiarto et al. (2022).

Admin diarahkan ke halaman /admin, sedangkan siswa ke /dashboard. Formulir pendaftaran memuat NIK, nama lengkap, tempat dan tanggal lahir, serta pilihan jurusan IPA atau IPS. NIK harus terdiri atas 16 digit dan bersifat unik. Siswa juga mengunggah KK dan Ijazah/SKL dalam format PDF, JPG, JPEG, atau PNG dengan ukuran maksimum 5 MB per berkas.

Setelah formulir dikirim, Laravel memvalidasi masukan dan menyimpan data pendaftar berstatus pending apabila memenuhi ketentuan. Dokumen disimpan secara privat dan dicatat dalam basis data untuk diproses layanan Python melalui OCR, validasi berbasis aturan, dan Isolation Forest.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

3.1.3 Implementasi Layanan Machine Learning

Layanan machine learning menggunakan Python sebagai worker yang terhubung dengan aplikasi web melalui basis data bersama. Worker memeriksa pendaftar berstatus pending setiap lima detik, mengubahnya menjadi processing, lalu menjalankan tiga tahap validasi. Penggunaan OCR untuk ekstraksi teks dibahas oleh Toha dan Triayudi (2022), sedangkan Isolation Forest untuk deteksi anomali dikaji oleh Waspada et al. (2020).

Tahap pertama menggunakan EasyOCR untuk mengekstraksi teks dan nilai confidence. PDF dikonversi menjadi citra menggunakan PyMuPDF dengan mengambil halaman pertama, sedangkan gambar diproses menggunakan PIL. Hasil ekstraksi digunakan untuk menghitung rata-rata confidence. Tahap kedua menerapkan validasi berbasis aturan terhadap keberadaan teks, rata-rata confidence minimal 0,2 pada KK dan Ijazah/SKL, kecocokan NIK, serta pencarian kata pertama nama pendaftar pada KK. Pencocokan NIK menormalisasi huruf “l” dan “i” menjadi angka 1 serta “o” menjadi angka 0. NIK yang tidak cocok menyebabkan pemeriksaan gagal, sedangkan nama yang tidak ditemukan hanya menghasilkan peringatan.

Tahap ketiga menggunakan Isolation Forest dengan empat fitur: rata-rata confidence OCR, confidence terendah, panjang teks hasil ekstraksi, dan jumlah peringatan validasi aturan. Model dilatih secara offline menggunakan skrip train_iforest.py dengan 350 sampel sintetis (300 normal dan 50 anomali) berdasarkan statistik sekitar 10 dokumen asli. Model disimpan menggunakan joblib sebagai model.pkl dengan konfigurasi n_estimators=100, max_samples=auto, contamination=0.15, dan random_state=42. Nilai contamination mendekati proporsi anomali dataset sebesar 14,3%.

Prediksi Isolation Forest digabungkan dengan hasil validasi aturan. Status auto_lolos diberikan apabila model memprediksi sampel normal dan seluruh aturan terpenuhi. Sebaliknya, deteksi anomali atau kegagalan aturan menghasilkan perlu_review. Jika model tidak tersedia, sistem menggunakan validasi aturan sebagai mekanisme alternatif (fallback). Kedua status merupakan rekomendasi pemeriksaan, sedangkan keputusan akhir penerimaan atau penolakan tetap ditetapkan oleh panitia PPDB.

3.1.4 Implementasi Basis Data

Basis data bersama digunakan oleh aplikasi web dan layanan machine learning untuk menyimpan informasi akun, data pendaftar, dokumen, serta status validasi. Struktur utamanya terdiri atas tiga tabel, yaitu users untuk akun siswa dan admin, pendaftars untuk identitas dan hasil pemeriksaan pendaftar, serta dokumens untuk informasi berkas persyaratan dan hasil ekstraksi OCR. Rincian data dan fungsi masing-masing tabel disajikan pada Tabel 3.

Tabel 3. Struktur Tabel Utama Basis Data

Nama Tabel	Data/Kolom yang Disediakan	Fungsi
users	Nama, email, kata sandi, dan peran pengguna	Mengelola akun siswa dan admin
pendaftars	NIK, nama lengkap, tempat dan tanggal lahir, jurusan, status validasi, catatan pemeriksaan, dan skor Isolation Forest	Menyimpan data pendaftar serta informasi hasil validasi
dokumens	Jenis dokumen, lokasi berkas, teks hasil OCR, dan nilai confidence OCR	Mengelola dokumen persyaratan yang diunggah pendaftar

Relasi antartabel menggunakan primary key dan foreign key. Tabel pendaftars terhubung dengan users melalui user_id, sedangkan tabel dokumens terhubung dengan pendaftars melalui pendaftar_id. Struktur relasional tersebut sejalan dengan konsep pemodelan basis

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

data menggunakan Entity Relationship Diagram (ERD) yang dibahas oleh Mafirah et al. (2026).

Kolom status_validasi memiliki tujuh nilai, yaitu draft, pending, processing, auto_lolos, perlu_review, diterima, dan ditolak. Worker memperbarui status auto_lolos atau perlu_review, sedangkan admin menetapkan status akhir diterima atau ditolak. Dalam implementasi ini, kolom ocr_text, ocr_confidence, dan iforest_score telah tersedia, tetapi belum diisi oleh worker. Worker hanya memperbarui status_validasi dan rule_notes, sehingga hasil ekstraksi, nilai confidence, dan skor anomali belum tersimpan pada kolom yang disediakan.

3.1.5 Implementasi Antarmuka Sistem

Antarmuka sistem PPDB online dikembangkan untuk mendukung aktivitas siswa dan panitia dalam proses pendaftaran serta pemeriksaan dokumen. Halaman siswa mencakup dashboard, formulir pendaftaran, pemantauan status dokumen, dan jadwal seleksi. Sementara itu, halaman admin digunakan untuk mengelola data pendaftar, meninjau dokumen, menetapkan keputusan akhir, dan mengelola jadwal seleksi. Rangkuman halaman antarmuka, alamat akses, dan fungsinya disajikan pada Tabel 4.

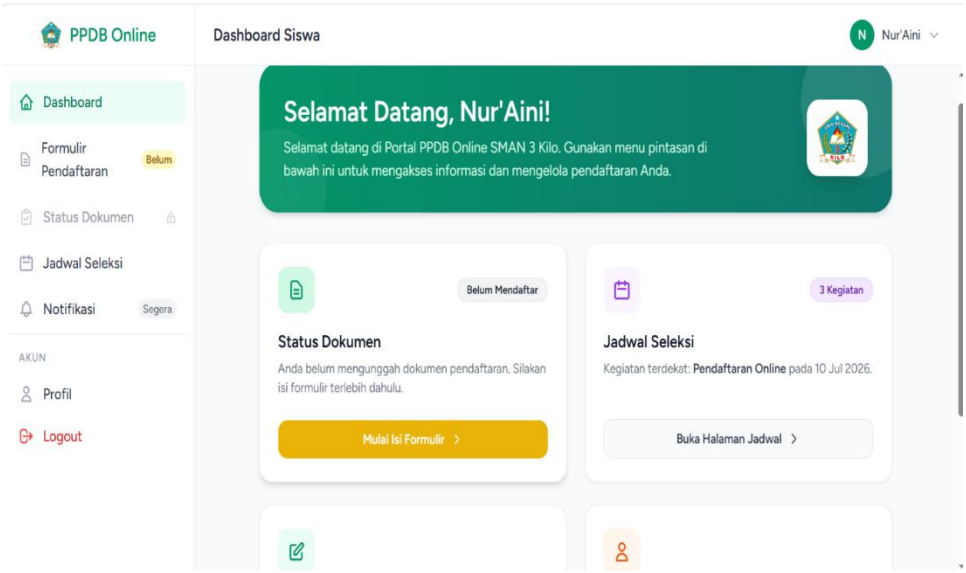
Tabel 4. Halaman Antarmuka Sistem

Halaman	URL	Akses	Fungsi
Landing page	/	Publik	Menampilkan informasi umum PPDB SMAN 3 Kilo
Registrasi	/register	Publik	Pembuatan akun siswa
Login	/login	Publik	Autentikasi siswa dan admin
Dashboard siswa	/dashboard	Siswa	Menyediakan pintasan ke status, jadwal, formulir, dan profil
Form pendaftaran	/pendaftaran	Siswa	Mengisi data diri dan mengunggah dokumen persyaratan
Status dokumen	/status-dokumen	Siswa	Memantau status validasi dan catatan sistem
Jadwal seleksi	/jadwal-seleksi	Siswa	Melihat jadwal seleksi yang dikelola admin
Dashboard admin	/admin	Admin	Menampilkan daftar pendaftar beserta statusnya
Detail pendaftar	/admin/pendaftar/{id}	Admin	Meninjau dokumen, catatan validasi, dan menetapkan keputusan

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

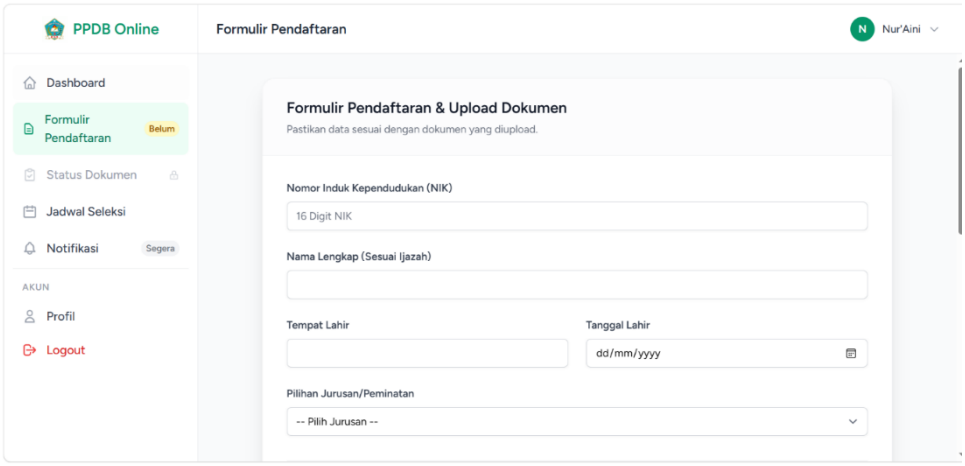
Jadwal admin	/admin/jadwal-seleksi	Admin	Menginput dan mengelola jadwal seleksi

Tabel 4 menunjukkan pembagian fungsi antarmuka berdasarkan hak akses pengguna. Siswa dapat melakukan pendaftaran dan mengakses informasi seleksi, sedangkan admin bertugas mengelola data pendaftar serta memeriksa dokumen.
Implementasi antarmuka siswa diawali dengan dashboard yang menjadi halaman utama setelah pengguna berhasil masuk ke sistem. Tampilan dashboard siswa disajikan pada Gambar 2.



Gambar 2. Tampilan Dashboard Siswa

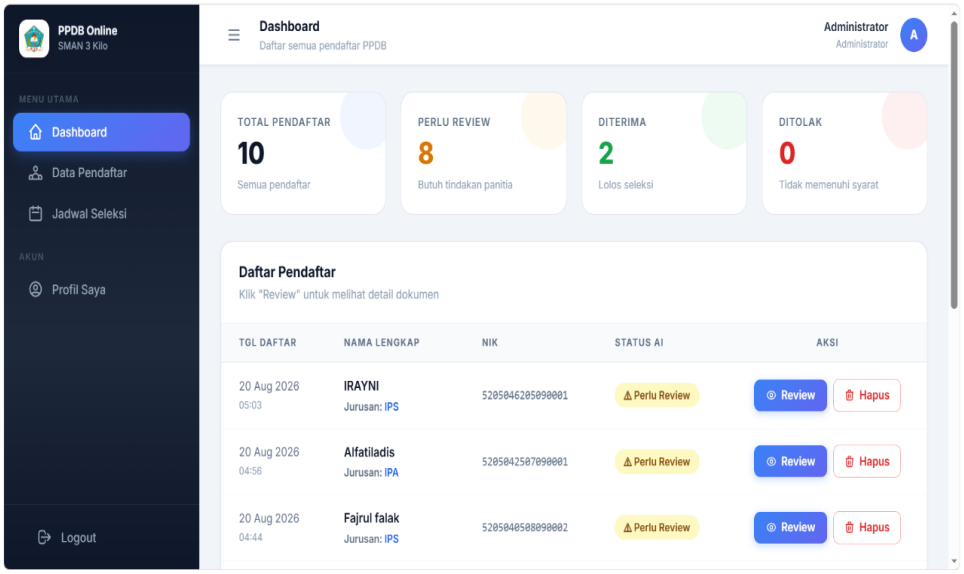
Gambar 2 menampilkan dashboard siswa dengan navigasi di sisi kiri serta banner sambutan dan kartu Status Dokumen dan Jadwal Seleksi pada area utama. Sebelum formulir pendaftaran dikirim, kartu Status Dokumen menampilkan label Belum Mendaftar dan tombol Mulai Isi Formulir, sedangkan menu Status Dokumen pada sidebar masih terkunci. Tampilan ini mengarahkan siswa untuk memulai proses pendaftaran.
Pengisian data diri dan pengunggahan dokumen persyaratan dilakukan melalui halaman formulir pendaftaran. Tampilan formulir tersebut ditunjukkan pada Gambar 3.



Gambar 3. Tampilan Form Pendaftaran

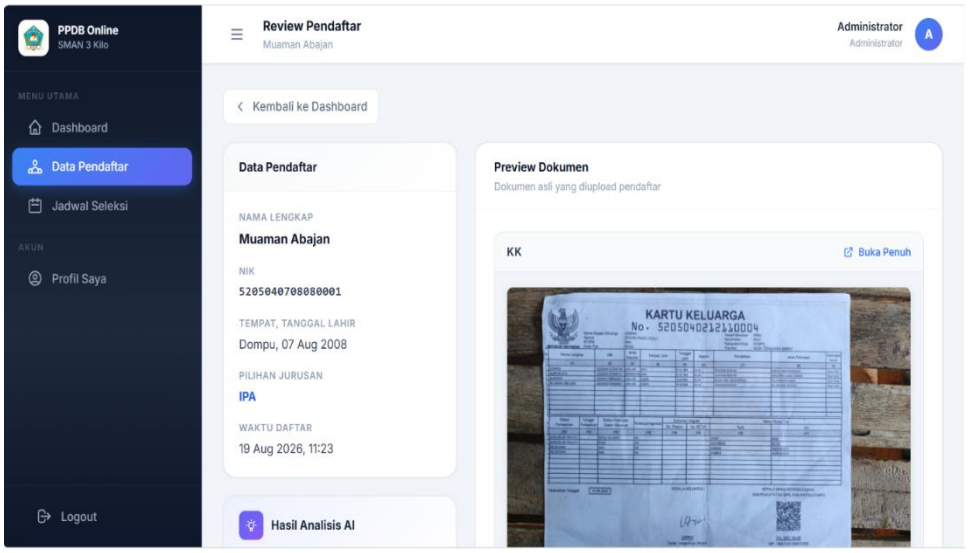
JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

Gambar 3 memperlihatkan halaman Formulir Pendaftaran dan Upload Dokumen. Siswa mengisi NIK, nama lengkap sesuai ijazah, tempat dan tanggal lahir, serta pilihan jurusan. Siswa juga mengunggah berkas persyaratan pendaftaran. Data dan dokumen yang dikirimkan selanjutnya disimpan oleh sistem untuk diproses melalui tahapan validasi.
Pada sisi pengelola, admin menggunakan dashboard untuk memantau data pendaftar dan status hasil validasi. Tampilan dashboard admin disajikan pada Gambar 4.



Gambar 4. Tampilan Dashboard Admin

Gambar 4 menunjukkan dashboard admin yang memuat kartu statistik jumlah total pendaftar, pendaftar yang perlu ditinjau, diterima, dan ditolak. Tabel Daftar Pendaftar menampilkan tanggal pendaftaran, nama, NIK, dan status hasil validasi AI. Tersedia tombol Review untuk membuka detail pendaftar dan tombol Hapus untuk menghapus data. Informasi tersebut membantu admin mengidentifikasi pendaftar yang memerlukan pemeriksaan lebih lanjut.
Pemeriksaan lebih lanjut dilakukan melalui halaman detail pendaftar yang menyediakan hasil analisis otomatis dan pratinjau dokumen yang diunggah siswa. Tampilan detail pendaftar pada sisi admin ditunjukkan pada Gambar 5.



Gambar 5. Tampilan Detail Pendaftar pada Sisi Admin

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

Gambar 5 menampilkan halaman Review Pendaftar pada sisi admin. Kolom kiri memuat rekomendasi dan catatan hasil analisis AI serta bagian Keputusan Final Panitia. Kolom kanan menampilkan pratinjau Kartu Keluarga dan Ijazah/SKL yang dapat dibuka dalam ukuran penuh. Informasi pribadi pendaftar pada tangkapan layar telah disamarkan. Tampilan ini menunjukkan bahwa hasil validasi otomatis menjadi bahan pertimbangan admin dalam memeriksa dokumen dan menetapkan keputusan akhir penerimaan.

3.2 Pengujian dan Evaluasi Sistem

Pengujian dilakukan untuk mengevaluasi fungsi aplikasi PPDB online dan mekanisme validasi dokumen menggunakan Optical Character Recognition (OCR), validasi berbasis aturan (rule-based validation), serta algoritma Isolation Forest. Pengujian mencakup tiga aspek, yaitu pengujian fungsional menggunakan metode Black Box, pengujian pipeline validasi dokumen, dan evaluasi kinerja model Isolation Forest. Hasilnya digunakan untuk menilai kesesuaian fungsi sistem serta mengidentifikasi keterbatasan implementasi.

3.2.1 Pengujian Fungsional (Black Box)

Pengujian fungsional menggunakan metode Black Box Testing untuk mengevaluasi kesesuaian fungsi aplikasi dengan kebutuhan sistem. Pengujian mencakup autentikasi, pengisian formulir, validasi data dan dokumen, pemantauan status pendaftaran, pengelolaan data pendaftar, serta pengaturan jadwal seleksi. Setiap skenario dijalankan dengan membandingkan hasil yang diharapkan dan respons sistem. Rangkuman pengujian disajikan pada Tabel 5.

Tabel 5. Pengujian Fungsional (Black Box)

No.	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian yang Tercatat
1	Login sebagai siswa	Diarahkan ke halaman dashboard siswa	Sistem berhasil menampilkan halaman dashboard siswa
2	Login sebagai admin	Otomatis diarahkan ke halaman admin	Sistem menampilkan halaman login admin
3	Submit formulir lengkap dengan KK dan Ijazah/SKL yang valid	Rekaman pendaftar dibuat berstatus `pending` dan muncul pesan sukses	Sistem menampilkan halaman formulir
4	Submit formulir dengan NIK kurang dari 16 digit	Ditolak dengan pesan kesalahan validasi	Sistem berhasil mendeteksi NIK kurang dari 16 digit
5	Submit formulir dengan NIK yang sudah terdaftar	Ditolak karena NIK harus unik	Sistem mendeteksi NIK yang sudah terdaftar
6	Unggah berkas berformat tidak didukung atau lebih dari 5 MB	Ditolak dengan pesan kesalahan	Sistem berhasil mendeteksi dokumen yang lebih dari 5 MB

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

7	Siswa membuka halaman status dokumen	Tampil lencana status dan catatan sistem	Sistem menampilkan halaman status dokumen siswa
8	Admin membuka detail pendaftar	Data diri, pratinjau dokumen, dan catatan validasi tampil	Sistem menampilkan halaman detail pendaftar
9	Admin menetapkan keputusan Diterima atau Ditolak	Status pendaftar berubah sesuai keputusan	Sistem membaca ada dokumen yang buram dan perlu ditinjau ulang oleh admin
10	Admin menambah atau mengubah jadwal seleksi	Jadwal tersimpan dan tampil pada dashboard siswa	Sistem berhasil menambah atau mengubah jadwal seleksi

Hasil pengujian pada Tabel 5 menunjukkan bahwa sebagian besar fungsi utama sistem telah berjalan. Namun, terdapat tiga skenario yang belum sepenuhnya sesuai dengan hasil yang diharapkan, yaitu login admin yang masih menampilkan halaman login, pengiriman formulir yang belum memastikan data tersimpan dengan status pending, serta penetapan keputusan admin yang masih mencatat dokumen buram untuk ditinjau ulang. Ketiga skenario tersebut memerlukan verifikasi lebih lanjut.

3.2.2 Pengujian Pipeline Validasi

Pengujian pipeline dilakukan untuk mengamati pemrosesan dokumen KK dan Ijazah/SKL melalui tahapan OCR, validasi berbasis aturan, dan deteksi anomali menggunakan Isolation Forest. Pengujian menggunakan variasi dokumen jelas, buram, terpotong, tidak sesuai jenis persyaratan, serta dokumen dengan ketidaksesuaian data. Hasil pengujian diamati berdasarkan status validasi dan catatan pemeriksaan yang dihasilkan sistem, sebagaimana disajikan pada Tabel 6.

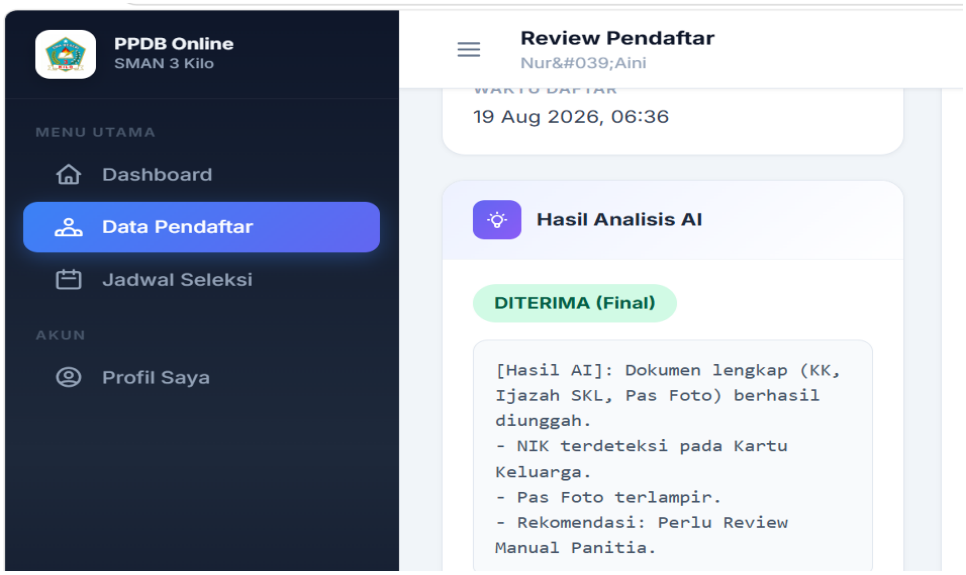
Tabel 6. Hasil Pengujian Pipeline Validasi Dokumen

No.	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian	Kesesuaian
1	KK dan Ijazah/SKL dalam kondisi jelas serta data sesuai	Status `auto_lolos`	Berhasil	Sesuai
2	Dokumen berupa foto buram	Status `perlu_review`	Berhasil	Sesuai
3	Dokumen dalam kondisi terpotong	Status `perlu_review`	Berhasil	Sesuai
4	NIK pada formulir berbeda dengan NIK pada KK	Status `perlu_review` karena tidak memenuhi validasi Layer 2	Berhasil	Sesuai

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

5	Berkas yang diunggah bukan dokumen KK atau Ijazah/SKL	Status `perlu_review`	Berhasil	Sesuai
6	Nama pada formulir berbeda dengan nama pada KK	Peringatan tercatat dalam `rule_notes`	Berhasil	Sesuai
7	Dokumen menggunakan format PDF	Halaman pertama PDF berhasil dikonversi dan teks dapat diproses oleh OCR	Berhasil	Sesuai

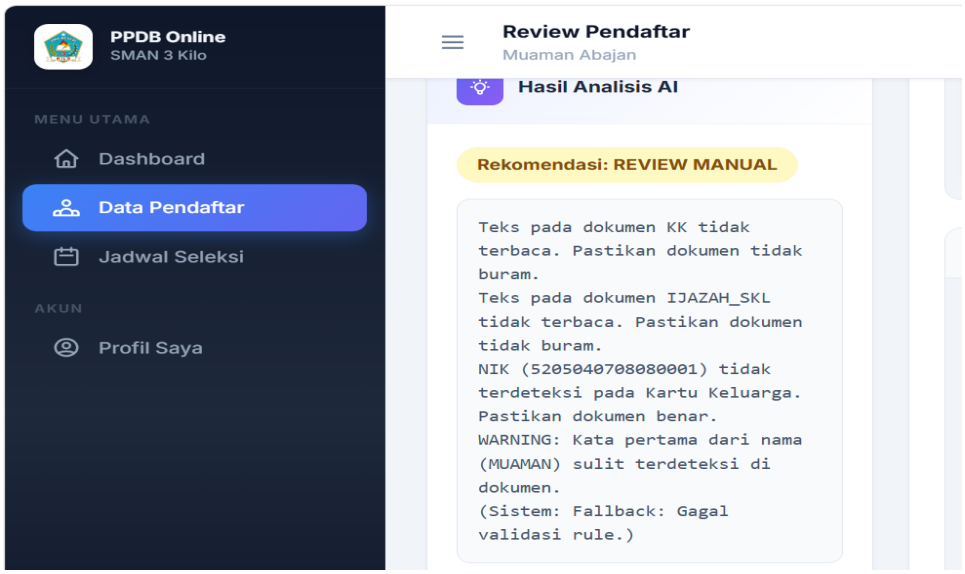
Berdasarkan Tabel 6, dokumen dengan kondisi jelas dan data sesuai memperoleh status auto_lolos. Sementara itu, dokumen buram, terpotong, bukan merupakan KK atau Ijazah/SKL, serta memiliki ketidaksesuaian NIK menghasilkan status perlu_review. Perbedaan nama antara formulir dan KK dicatat sebagai peringatan melalui rule_notes, sehingga dapat menjadi bahan pemeriksaan admin.
Hasil pemrosesan dokumen dengan kondisi jelas ditampilkan pada Gambar 6. Gambar tersebut memperlihatkan hasil validasi ketika dokumen dapat diproses dan memenuhi kriteria pemeriksaan otomatis.



Gambar 6. Hasil Validasi pada Dokumen yang Jelas

Gambar 6 memperlihatkan bahwa dokumen dengan kualitas visual memadai dapat diproses melalui OCR sehingga teksnya digunakan dalam pemeriksaan berbasis aturan dan deteksi anomali. Apabila hasil pemeriksaan memenuhi ketentuan, sistem memberikan status auto_lolos. Status ini menunjukkan terpenuhinya kriteria pemeriksaan otomatis, bukan keputusan akhir penerimaan.
Sebaliknya, dokumen dengan kualitas visual kurang baik dapat memengaruhi hasil ekstraksi teks dan pemeriksaan berikutnya. Kondisi tersebut ditunjukkan pada Gambar 7.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151



Gambar 7. Hasil Validasi pada Dokumen Buram atau Terpotong

Gambar 7 menunjukkan bahwa dokumen buram atau terpotong dapat memperoleh status perlu_review, yang menandakan perlunya pemeriksaan lebih lanjut oleh admin sebelum keputusan akhir ditetapkan. Dengan mekanisme tersebut, dokumen yang belum dapat dipastikan kesesuaiannya tidak langsung dinyatakan memenuhi persyaratan.
Secara keseluruhan, pengujian menunjukkan bahwa pipeline dapat memproses berbagai kondisi dokumen dan menghasilkan status sesuai skenario yang diuji. Integrasi OCR, validasi berbasis aturan, dan Isolation Forest mendukung pemeriksaan awal, sedangkan verifikasi akhir tetap dilakukan oleh admin untuk memastikan kesesuaian dokumen persyaratan.

3.2.3 Evaluasi Model Isolation Forest

Evaluasi Isolation Forest menggunakan 10 data uji, terdiri atas lima data normal dan lima anomali. Hasil prediksi dibandingkan dengan kategori aktual menggunakan confusion matrix pada Tabel 7.

Tabel 7. Confusion Matrix Model Isolation Forest

Kategori Aktual	Prediksi Normal	Prediksi Anomali
Normal	4	1
Anomali	1	4

Berdasarkan Tabel 7, model mengklasifikasikan empat data normal dan empat data anomali secara tepat, sedangkan masing-masing satu data mengalami kesalahan klasifikasi. Hasil pengukuran pada Tabel 8 menunjukkan nilai precision, recall, dan F1-score masing-masing sebesar 80%.

Tabel 8. Hasil Evaluasi Model Isolation Forest

Metrik Evaluasi	Nilai
Precision	80%
Recall	80%

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

F1-score	80%

Hasil tersebut perlu ditafsirkan secara terbatas karena model dilatih menggunakan 350 sampel sintetis (300 normal dan 50 anomali) berdasarkan statistik sekitar 10 dokumen asli. Penggunaan pola pembangkitan data yang sama pada pelatihan dan pengujian berpotensi menghasilkan penilaian yang lebih baik dibandingkan kinerja pada dokumen nyata yang lebih beragam. Oleh karena itu, nilai 80% merupakan hasil awal yang belum dapat digeneralisasikan. Pengujian dengan lebih banyak dokumen asli tetap diperlukan, sedangkan keputusan akhir kelayakan dokumen menjadi kewenangan admin PPDB.

3.3 Pembahasan

Sistem PPDB di SMAN 3 Kilo mengintegrasikan pendaftaran berbasis web dengan validasi dokumen otomatis. Pengembangannya menggunakan RAD sebagaimana penelitian Yudahana et al. (2023) dan Kiayi et al. (2025), dengan penambahan pemeriksaan dokumen untuk mendukung administrasi penerimaan peserta didik.

Arsitektur sistem memisahkan aplikasi Laravel dan layanan machine learning Python melalui basis data bersama. Worker menggunakan mekanisme polling setiap lima detik untuk memproses pendaftaran tanpa Redis atau message broker. Pemeriksaan mengintegrasikan OCR (Toha dan Triayudi, 2022; Darpito et al., 2025), validasi berbasis aturan, dan Isolation Forest (Waspada et al., 2020; Mutmainah dan Yustanti, 2024; Ohyver et al., 2025). Hasilnya berupa status auto_lolos atau perlu_review sebagai rekomendasi, sedangkan keputusan diterima atau ditolak tetap menjadi kewenangan panitia melalui prinsip human-in-the-loop.

Pengujian menunjukkan bahwa fungsi utama aplikasi telah tersedia, meskipun beberapa skenario masih memerlukan verifikasi. Pipeline validasi dapat membedakan dokumen yang memenuhi kriteria dan yang memerlukan pemeriksaan admin. Evaluasi Isolation Forest menghasilkan precision, recall, dan F1-score masing-masing 80% berdasarkan 10 data uji. Hasil tersebut belum dapat digeneralisasi karena model dilatih menggunakan 350 sampel sintetis berdasarkan sekitar 10 dokumen asli.

Keterbatasan sistem meliputi pemrosesan PDF yang hanya mengambil halaman pertama, ambang confidence OCR 0,2, pemeriksaan nama berdasarkan kata pertama, serta nilai contamination 0,15 yang belum diuji dengan lebih banyak dokumen nyata. Selain itu, kolom ocr_text, ocr_confidence, dan iforest_score belum diisi oleh worker. Pengembangan selanjutnya perlu memperluas pengujian dengan dokumen asli, meningkatkan pemeriksaan dokumen, dan melengkapi penyimpanan hasil validasi. Secara keseluruhan, integrasi ketiga metode mendukung pemeriksaan awal dokumen PPDB, tetapi keputusan akhir tetap memerlukan verifikasi panitia.

4. KESIMPULAN

Penelitian ini menghasilkan sistem PPDB berbasis web untuk SMAN 3 Kilo, Kabupaten Dompu, yang mendukung pendaftaran, pengelolaan data peserta, pengunggahan dokumen, dan pemantauan status validasi. Pemeriksaan dokumen KK dan Ijazah/SKL mengintegrasikan tiga lapisan, yaitu OCR menggunakan EasyOCR, validasi berbasis aturan untuk memeriksa keterbacaan, format NIK dan konsistensi data, serta Isolation Forest untuk mendeteksi anomali. Sistem menghasilkan status auto_lolos atau perlu_review sebagai rekomendasi pemeriksaan.

Evaluasi Isolation Forest menghasilkan precision, recall, dan F1-score masing-masing sebesar 80% pada 10 data uji. Hasil tersebut belum dapat digeneralisasi karena keterbatasan dokumen asli dan penggunaan data augmentasi serta sintetis, sehingga penerapan model masih merupakan studi kelayakan (feasibility study). Sistem menerapkan prinsip human-in-the-loop; hasil validasi otomatis hanya menjadi informasi pendukung, sedangkan keputusan akhir diterima atau ditolak tetap menjadi kewenangan panitia PPDB.



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8, Tahun 2026
ISSN 3025-0919 (media online)
Hal 2137-2151

REFERENCES

Budiarto, S., Hidayat, T., & Nugroho, S. (2022). Sistem informasi penerimaan mahasiswa baru dengan framework Laravel. Bulletin of Information Technology (BIT), 3(2), 121–127. https://doi.org/10.47065/bit.v3i2.284

Darpito, M. N., Firdausy, K., & Fadlil, A. (2025). Perbandingan unjuk kerja library optical character recognition (OCR) dalam pengenalan teks pada dokumen digital. Jurnal Informatika Polinema, 11(3), 273–282. https://doi.org/10.33795/jip.v11i3.7025

Fedhira, & Prianto, C. (2025). Systematic literature review: Analysis of AI implementation for document verification. Jurnal Ilmiah Sistem Informasi, 4(3), 417–430. https://doi.org/10.51903/kjjwk708

Kiayi, M. Z. I., Rantung, V. P., & Rorimpandey, G. C. (2025). Aplikasi pendaftaran siswa baru berbasis web menggunakan metode rapid application development (RAD). Journal of Informatics, Business, Education and Innovation Technology, 3(2), 83–97.

Mafirah, D., Hary, A., Sugiran, S., & Annisa, L. (2026). Evaluasi strategi pemodelan basis data melalui entity relationship diagram (ERD) dalam perspektif akademik dan praktik industri: Literature review. Jurnal Pendidikan Tambusai, 10(1), 3935–3943. https://doi.org/10.31004/jptam.v10i1.36705

Mutmainah, M., & Yustanti, W. (2024). Studi komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada analisis anomali kinerja dosen. Journal of Informatics and Computer Science (JINACS), 6(02), 532–540. https://doi.org/10.26740/jinacs.v6n02.p532-540

Ohyver, G. A., Beng, J. T., Trisnawarman, D., Tiatri, S., & Ohyver, D. A. (2025). Analisis data keuangan menggunakan Isolation Forest untuk meningkatkan akurasi deteksi anomali. INTECOMS: Journal of Information Technology and Computer Science, 8(6), 1854–1860. https://doi.org/10.31539/zmbdyd52

Rosa, A. S., & Shalahuddin, M. (2024). Rekayasa perangkat lunak terstruktur dan berorientasi objek. Informatika.

Toha, M. R., & Triayudi, A. (2022). Penerapan membaca tulisan di dalam gambar menggunakan metode OCR berbasis website pada e-KTP. Jurnal Sains dan Teknologi, 11(1), 175–183.

Waspada, I., Bahtiar, N., Wirawan, P. W., & Awan, B. D. A. (2020). Performance analysis of isolation forest algorithm in fraud detection of credit card transactions. Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika, 6(2). https://doi.org/10.23917/khif.v6i2.10520

Yudahana, A., Riadi, I., & Elvina, A. (2023). Perancangan sistem informasi pendaftaran peserta didik baru (PPDB) berbasis web menggunakan metode rapid application development (RAD). RABIT: Jurnal Teknologi dan Sistem Informasi Univrab, 8(1), 47–58.