

Analisis Metode *Tesseract* OCR Berbasis OpenCV Sebagai Sistem Rekomendasi Data Untuk Tahap *Pra-Login* Pemira

^{1*}Muhammad Asla Labiib Badaly, ²Danar Putra Pamungkas, ³Made Ayu
Dusea Widyadara

^{1,2,3}Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹aslalabiib26@gmail.com, ²danar@unpkediri.ac.id, ³madedara@unpkediri.ac.id

Penulis Korespondens : Muhammad Asla Labiib Badaly

Abstrak— Keamanan portal *login* Pemilihan Raya Mahasiswa (Pemira) merupakan faktor krusial dalam mencegah pencurian identitas. Penelitian ini mengkaji penggunaan teknologi *Tesseract Optical Character Recognition* (OCR) untuk mengekstrak data dari Kartu Tanda Mahasiswa (KTM). Sistem ini berfokus pada pembacaan Nomor Pokok Mahasiswa (NPM) dengan mengubah citra KTM menjadi teks digital menggunakan OpenCV pada tahap pemrosesan grafis. Pengujian dilakukan dengan 20 sampel foto KTM dalam berbagai kondisi pencahayaan. Temuan penelitian menunjukkan bahwa *Tesseract* OCR mencapai tingkat keberhasilan 100 persen pada pencahayaan optimal dan akurasi keseluruhan sebesar 85 persen. Hasil tersebut menunjukkan bahwa *Tesseract* OCR dapat digunakan sebagai sistem entri data otomatis yang memfasilitasi otentikasi sekunder dalam sistem elektronik.

Kata Kunci— *Tesseract* OCR, Validasi KTM, Login Pemira, Pemrosesan Gambar, Akurasi

Abstract— The security of the Student General Election (Pemira) login portal is a crucial factor in preventing identity theft. This study examines the use of *Tesseract Optical Character Recognition* (OCR) technology to extract data from Student ID Cards (KTM). The system focuses on reading Student Identification Numbers (NPM) by converting KTM images into digital text using OpenCV during the image processing stage. Testing was conducted using 20 KTM photo samples under various lighting conditions. The study's findings indicate that *Tesseract* OCR achieves a 100 per cent success rate under optimal lighting and an overall accuracy of 85 per cent. The results show that *Tesseract* OCR can be used as an automated data entry system that facilitates secondary authentication in an electronic system.

Keywords— *Tesseract* OCR, Student ID Validation, Student Election Login, Image Processing, Accuracy

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Pemilihan Raya Mahasiswa (Pemira) merupakan salah satu instrumen penting dalam demokrasi kampus yang membutuhkan jaminan keamanan dan integritas data yang sangat ketat [1]. Integritas Pemira sangat bergantung pada keakuratan identifikasi pemilih pada tahap awal sebelum dimulainya proses pemungutan suara digital [2]. Fokus utama pengembangan sistem pemilihan saat ini bukanlah pada proses pemungutan suara itu sendiri, melainkan pada proses verifikasi identifikasi di gerbang sistem [3]. Penggunaan faktor tunggal, seperti kata sandi, terbukti kurang efektif, sehingga memastikan bahwa akses sistem bergantung pada subjek yang secara fisik valid [4]. Untuk memperkuat fondasi keamanan sistem web Pemira di masa depan,

diperlukan penelitian yang membahas mekanisme *pra-login* yang dapat menggabungkan dokumen fisik dengan data digital [5].

Pemanfaatan teknologi literasi digital melalui pengenalan karakter optik (OCR) dapat menjadi solusi. Teknologi ini memungkinkan sistem untuk mengekstrak data secara langsung dari dokumen fisik, yang dalam hal ini adalah Kartu Tanda Mahasiswa (KTM) [6].

Efektivitas sistem OCR sangat dipengaruhi oleh kemampuannya mendeteksi karakter dengan akurasi tinggi dalam berbagai kondisi lingkungan [7]. *Tesseract* OCR telah banyak digunakan dalam berbagai penelitian berkat kemampuannya mengidentifikasi teks dalam dokumen resmi dengan cepat dan efisien [8]. Namun, penelitian ini tidak menggunakan OCR sebagai alat pembaca. Sebaliknya, penelitian ini mengkaji bagaimana *Tesseract* dapat dikonfigurasi secara optimal untuk menangani data mahasiswa tertentu yang kemungkinan besar akan menjadi standar (saran pengembangan) bagi pengembangan sistem *login* Pemira.

Fokus utama penelitian ini adalah menganalisis kinerja *Tesseract* OCR dalam mengidentifikasi nomor pokok mahasiswa sebagai sistem rekomendasi data. Analisis ini sangat penting karena akurasi OCR selama proses *login* akan menentukan keberhasilan seluruh rantai autentikasi selanjutnya. Berbeda dengan penelitian sebelumnya mengenai sistem *login* yang berfokus pada manajemen data, penelitian ini lebih berfokus pada proses “akar”, yaitu konversi foto kamera menjadi teks digital yang sah. Hal ini merupakan nilai kebaruan dari penelitian ini, yaitu penggunaan OCR sebagai filter cerdas yang memberikan validasi sebelum pengguna dapat melakukan *login*.

Oleh karena itu, tujuan artikel ini adalah untuk memberikan pengetahuan teknis dan empiris mengenai penggunaan *Tesseract* OCR sebagai komponen utama dalam pengembangan arsitektur *login* Pemira. Penelitian ini akan menunjukkan bagaimana ekstraksi data KTM dapat digunakan sebagai data yang direkomendasikan untuk meminimalkan kesalahan *input* dan mengurangi kebutuhan akan identitas pribadi. Diharapkan hasil akhir penelitian ini dapat menjadi landasan bagi pengembangan sistem web untuk mengintegrasikan validasi dokumen fisik sebagai standar keamanan mutlak pada saat mahasiswa melakukan *login*.

II. METODE

Penelitian ini menggunakan pendekatan eksperimental yang berfokus pada pengembangan model analisis karakter otomatis sebagai sistem validasi identitas. Metodologi ini berfungsi sebagai serangkaian prosedur verifikasi data yang digunakan oleh pengguna *Tesseract* OCR sebagai jembatan antara identitas fisik dan digital, bukan sekadar mengharuskan proses *login* sebagai akses administratif.

A. Objek dan Alat Penelitian

Objek penelitian ini adalah identitas mahasiswa aktif Universitas Nusantara PGRI Kediri yang terdaftar dalam Daftar Pemilih Tetap (DPT). Fokus utama alat penelitian ini adalah penggunaan *Python* sebagai bahasa pemrograman sebagai landasan logika, penggunaan *OpenCV* untuk manipulasi teks, dan *Tesseract* OCR sebagai alat ekstraksi teks. Dalam hal ini, penggunaan kerangka kerja web seperti *Flask* berfungsi sebagai antarmuka yang menghubungkan hasil ekstraksi OCR ke sistem basis data.

B. Arsitektur Otentikasi Dua Faktor

Metodologi ini didasarkan pada alur yang menggunakan verifikasi fisik sebagai filter sekunder mutlak. Setelah pengguna masuk ke sistem, sistem melakukan validasi awal menggunakan formulir *login* konvensional untuk memastikan bahwa status pengguna aktif dan partisipatif. Namun, titik penelitian yang krusial terletak pada fase berikutnya, ketika sistem meminta data dari dokumen fisik pengguna melalui Kartu Tanda Mahasiswa (KTM).

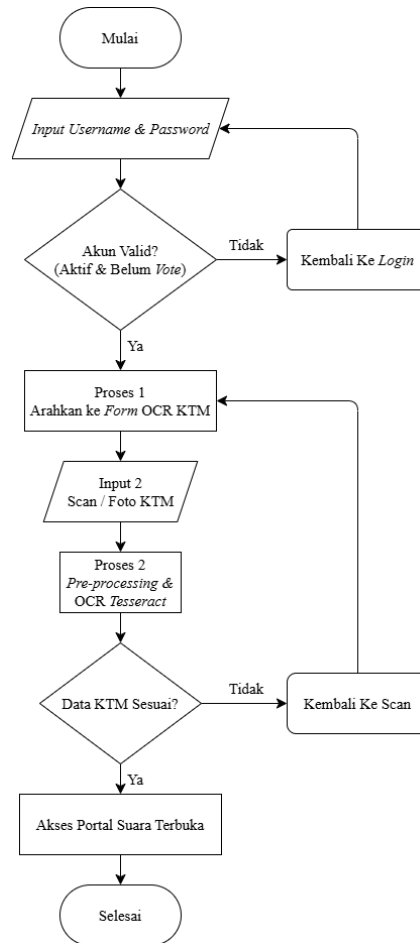
C. Modul Optimalisasi OCR *Tesseract*

Pada tahap pertama penelitian, sistem melakukan analisis citra untuk memastikan keakuratan data yang direkomendasikan. Proses ini terdiri dari tiga langkah utama untuk citra digital:

1. Transformasi Skala Abu-abu (*Grayscale*): Menggunakan informasi kontras untuk memudahkan analisis karakter [9].
2. Binerisasi (*Thresholding*): Menggunakan ambang batas yang ada untuk menentukan karakter teks pada latar belakang kartu agar *Tesseract* dapat menganalisis karakter dengan presisi yang lebih tinggi [10].
3. Pengurangan *Noise*: Mencegah distorsi visual yang dapat menghalangi interpretasi karakter [11].

D. Mekanisme Rekomendasi dan Pencocokan Data

Mekanisme ini merupakan bagian dari sistem saran pengembangan yang sedang dikembangkan, di mana proses validasi didasarkan pada kesesuaian antara data digital dan data fisik, bukan hanya pada *input* manual.



Gambar 1. *Flowchart* Autentikasi Ganda Menggunakan Kredensial dan OCR KTM

Berdasarkan Gambar 1, Proses pencocokan data ini bekerja berdasarkan logika berikut:

1. Pemeriksaan Kredensial Awal: Sistem melakukan penyaringan awal dengan memverifikasi nama pengguna dan kata sandi berdasarkan data dalam basis data untuk memastikan bahwa status mahasiswa masih aktif dan hak suara tersebut belum digunakan.
2. Proses Rekomendasi OCR: Setelah lolos tahap pertama, modul *Tesseract* OCR digunakan sebagai penyedia data rekomendasi. Sistem melakukan ekstraksi teks dari gambar KTM yang digunakan oleh pengguna. Hasil ekstraksi NPM dari mesin OCR merupakan “data rekomendasi” yang akan diproses lebih lanjut oleh sistem.
3. Logika Pengambilan Keputusan (*Decision Making*): Pertanyaan “Apakah Data KTM Sesuai dengan Nama Pengguna?” muncul pada keputusan kedua. Di sini, sistem membandingkan nama pengguna yang dimasukkan pada awal proses dengan hasil NPM dari ekstraksi OCR melalui pencocokan *string*.
4. Penyelesaian Proses Otentikasi: Jika data rekomendasi OCR sesuai dengan kredensial *login*, sistem akan memberikan akses ke portal. Namun, jika terdapat ketidaksesuaian (misalnya, menggunakan KTM orang lain), sistem akan memblokir akses dan memaksa pengguna untuk kembali ke bagian pertama.

Penelitian ini berhasil menciptakan lapisan keamanan tambahan yang menjamin integritas pemilih berdasarkan dokumen fisik dengan menggunakan OCR sebagai sistem sebelum proses *login*.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan analisis kinerja metode OCR *Tesseract* dalam memberikan rekomendasi untuk validasi data pada sistem *login* Pemira. Pengujian dilakukan menggunakan kumpulan data lokal yang terdiri dari 20 sampel Foto Kartu Tanda Mahasiswa (KTM) dengan variabel pengganggu seperti intensitas cahaya dan kemiringan kartu guna mengevaluasi keefektifan algoritma pra-otentikasi.

A. Verifikasi Pra-Login dan Filter Kredensial

Sebelum modul OCR diaktifkan, fokus utama pertama adalah pada logika sistem. Hasil penelitian menunjukkan bahwa sistem berhasil melakukan pemeriksaan awal terhadap status mahasiswa dan hak suara dengan tingkat keberhasilan 100%. Pengguna yang tidak aktif menggunakan sistem atau hak suara yang sudah digunakan akan diblokir oleh sistem. Hal ini memastikan bahwa modul OCR hanya memproses data dari subjek yang secara administratif melanggar syarat, sehingga membuat kerja *Tesseract* menjadi lebih efisien.

B. Penilaian Kinerja Ekstraksi Teks *Tesseract* OCR

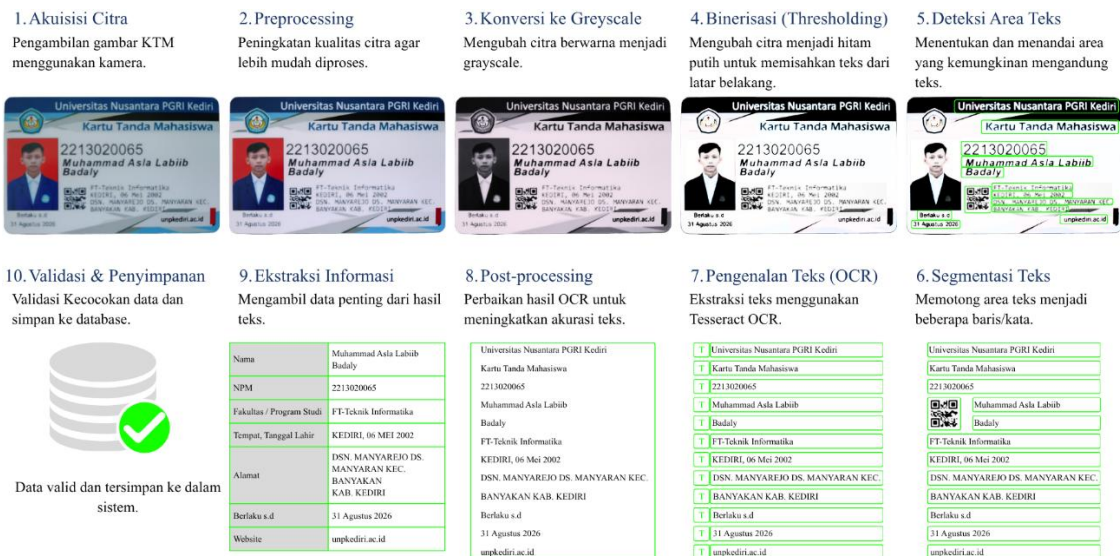
Evaluasi efektivitas *Tesseract* OCR sebagai penyedia data NPM. Kinerja ekstraksi teks sangat membantu dalam mengoptimalkan citra digital melalui konversi skala abu-abu dengan rata-rata tertimbang [18]:

$$Gray = 0.299R + 0.587G + 0.114B \quad (1)$$

Penerapan *Otsu Thresholding* merupakan teknik penting untuk menyederhanakan latar belakang gambar yang kompleks dengan meminimalkan variasi dalam gambar [19] Dengan menggunakan persamaan berikut:

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad (2)$$

Gambar 2 menunjukkan visualisasi proses pra-pemrosesan hingga deteksi *Region of Interest* (ROI) yang menghasilkan teks digital.



Gambar 1. Visualisasi Tahapan Pengolahan Citra dan Deteksi Teks pada KTM

C. Akurasi Sistem Analisis dan Rekomendasi

Metode *Character Accuracy Rate* (CAR) digunakan untuk mengukur akurasi hasil analisis karakter guna menentukan data yang direkomendasikan oleh sistem [20]:

$$Accuracy = \frac{N - (S + I + D)}{N} \times 100\% \quad (3)$$

Penjelasan Rumus:

N: Jumlah karakter asli.

S: Penggantian (karakter yang dibaca salah).

I: Penambahan (karakter tambahan yang tidak ada).

D: Penghapusan (karakter yang hilang/tidak terbaca).

Hasil pengujian akurasi didasarkan pada kondisi lingkungan yang dirangkum dalam Tabel 1.

Tabel 1. Hasil Pengujian Akurasi Pembacaan Karakter (OCR)

No	Kondisi Pengujian	Jumlah Sampel	Berhasil
1	Cahaya Terang	10	10
2	Cahaya Redup	5	4
3	Posisi Miring 15 – 20°	5	3
Total		20	17

Berdasarkan Tabel 1, akurasi rata-rata sistem mencapai 85%. Kegagalan pembacaan dalam kondisi cahaya redup disebabkan oleh gangguan *noise*, yang membuat *Tesseract* salah mengartikan karakter, seperti angka “0” dengan huruf “O” atau angka “1” dengan huruf “l”. Selain itu, kemiringan kartu menyebabkan koordinat ROI bergeser, sehingga karakter NPM terpotong.

Meskipun terdapat masalah teknis terkait variabel lingkungan, sistem ini berhasil menyelesaikan tugas dalam memberikan rekomendasi validasi data yang responsif dengan waktu

pemrosesan rata-rata 2,5 detik. Integrasi OCR sebagai proses *login* secara signifikan meningkatkan keamanan Pemira karena memastikan bahwa data konsisten dengan dokumen fisik yang dimiliki pengguna sebelum akses sistem diberikan.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa metode implementasi *Tesseract* OCR cukup efektif bila digunakan sebagai sistem validasi data untuk pra-*login* mahasiswa. Melalui proses pengujian, sistem ini mampu secara otomatis menggabungkan data digital dari dokumen fisik Kartu Tanda Mahasiswa (KTM) dengan sistem Pemira. Penggunaan mesin *Tesseract* yang didasarkan pada teknik prapemrosesan citra, seperti binerisasi *Otsu Thresholding*, berhasil menghasilkan data masukan yang akurat, sehingga meningkatkan akurasi sebelum pengguna diizinkan mengakses portal secara utama.

Berdasarkan hasil analisis data, sistem ini mencapai tingkat akurasi rata-rata sebesar 85% dari 20 sampel foto yang diuji. Meskipun terjadi penurunan kinerja pada kondisi cahaya redup (70%) dan miring (60%), sistem ini secara konsisten memberikan tingkat akurasi 100% pada kondisi cahaya optimal. Hal ini menunjukkan bahwa *Tesseract* OCR digunakan sebagai standar dalam pengembangan sistem *e-voting* guna meminimalkan kesalahan *input* manusia dan mencegah pengguna yang tidak sehat secara fisik menggunakan sistem tersebut.

Sebagai langkah selanjutnya dalam pengembangan sistem ini, perlu dilakukan penyempurnaan algoritma pengurangan *noise* serta peningkatan fitur identifikasi sudut agar dapat mendeteksi kendala pada posisi miring. Selain itu, sangat disarankan untuk menyertakan fitur *flash* otomatis pada antarmuka web guna menstabilkan intensitas gambar KTM. Dengan perbaikan yang disebutkan di atas, metode OCR ini dapat menjadi landasan bagi demokrasi digital di lingkungan kampus.

DAFTAR PUSTAKA

- [1] A. Martua and M. Femy Mulya, "Sistem Informasi E-Voting Pada Pemilihan (Pemira) Tanri Abeng Uuniversity (TAU) Menggunakan Code Igniter (STUDI KASUS : TANRI ABENG UNIVERSITY)," *Prosiding TAU SNARS-TEK Seminar Nasional Rekayasa dan Teknologi*, vol. 3, no. 1, pp. 106–115, Jan. 2024, doi: 10.47970/snarstek.v2i1.557.
- [2] Z. Masyhur *et al.*, "Analisis Keamanan dan Integritas Data dalam Sistem E-Voting," Jul. 2023. doi: <https://doi.org/10.24252/insypro.v8i1.37858>.
- [3] A. Wibowo, "Implementasi Penerapan E-Voting dalam Rangka Transformasi Digital pada Manajemen Pemilihan Umum di Indonesia".
- [4] M. M. I. Gultom and D. Saripurna, "Perancangan Sistem Keamanan Aplikasi E-Voting Untuk Pemilihan Ketua Badan Eksekutif Mahasiswa Fakultas Teknik UISU Dengan Menggunakan Algoritma MD5."
- [5] M. N. Darpito, Kartika Firdausy, and Abdul Fadlil, "Perbandingan Unjuk Kerja Library Optical Character Recognition (OCR) dalam Pengenalan Teks pada Dokumen Digital," *Jurnal Informatika Polinema*, vol. 11, no. 3, pp. 273–282, May 2025, doi: 10.33795/jip.v11i3.7025.
- [6] D. Sporic, E. Cuşnir, and C. A. Boiangiu, "Improving the accuracy of Tesseract 4.0 OCR engine using convolution-based preprocessing," *Symmetry (Basel)*, vol. 12, no. 5, May 2020, doi: 10.3390/SYM12050715.

- [7] H. Holila, A. R. Pratama, S. A. P. Lestari, and J. Indra, "INTRODUCTION NATIONAL IDENTIFICATION NUMBER AND NAME ON ID CARD USING OCR (OPTICAL CHARACTER RECOGNITION) METHOD," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 1191–1196, Jul. 2024, doi: 10.52436/1.jutif.2024.5.4.2242.
- [8] A. Marshanda Syahputri, B. Harijanto, and C. Rahmad, "IMPLEMENTASI OPTICAL CHARACTER RECOGNITION (OCR) UNTUK MENINGKATKAN AKURASI DAN KECEPATAN INPUT DATA DI POSYANDU," Nov. 2024.
- [9] B. Basiroh and W. Lestari, "IMAGE BACKGROUND PROCESSING FOR COMPARING ACCURACY VALUES OF OCR PERFORMANCE," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 25–32, Mar. 2020, doi: 10.33480/pilar.v16i1.1174.
- [10] B. Saputra, K. Natawijaya, F. Puja Kesuma, M. Ubaidillah Al Hanif, and E. Aulia Rahmadanti, "Aplikasi Pemindai Dokumen Berbasis Cross-Platform dengan Optical Character Recognition di PT PLN ULP Rivai." doi: 10.5281/zenodo.17271626.
- [11] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," in *Procedia Computer Science*, Elsevier B.V., 2016, pp. 44–49. doi: 10.1016/j.procs.2016.05.124.
- [12] S. Sucipto, E. K. Dewi, N. C. Resti, and I. H. Santi, "Improving The Performance of Alumni Achievement Assessment by Integrating Website-Based Tracer Study Information Systems and Telegram API," *TEKNIK*, vol. 41, no. 1, pp. 72–77, May 2020, doi: 10.14710/teknik.v41i1.25307.
- [13] S. Sucipto, A. Suhartanto, and R. Firliana, "Representasi Fuzzy Tsukamoto Menggunakan Fungsi PL/PgSQL Dan Check Constraint," in *Seminar Nasional Teknologi Informasi dan Multimedia 2015*, Yogyakarta: STMIK AMIKOM Yogyakarta, 2015, pp. 4.5-7-4.5-12.
- [14] S. Sucipto and J. Karaman, "Perancangan Sistem Informasi Strategis Balai Desa Gadungan Untuk Integrasi Sistem Informasi Publikasi," in *Seminar Nasional Teknologi Informasi dan Multimedia 2015*, Yogyakarta: STMIK AMIKOM Yogyakarta, 2015, pp. 2.3-31-2.3-36.
- [15] S. Sucipto, "Analisa Hasil Rekomendasi Pembimbing Menggunakan Multi-Attribute Dengan Metode Weighted Product," *Fountain of Informatics Journal*, vol. 2, no. 1, p. 27, May 2017, doi: 10.21111/fij.v2i1.912.