

Implementasi Algoritma OCR (*Optical Character Recognition*) untuk Ekstraksi Informasi dari Citra Struk Transaksi Keuangan

Latifa Khoirani^{1*}, Rino Ariansyah², Supiyandi³

Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi,

^{1,2}Universitas Islam Negeri Sumatera Utara, Indonesia

³Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: khoiranilatifa@gmail.com¹, rino.ariansyah16@gmail.com²,

supiyandi.mkom@gmail.com³

Korespondensi penulis: khoiranilatifa@gmail.com*

Abstract : *This research aims to automate the process of extracting information from financial transaction receipts using Optical Character Recognition (OCR) algorithm. The OCR algorithm is used to recognize characters from digital images of transaction receipts so that information such as date, transaction amount, and other details can be accurately identified. By applying image processing methods, this research successfully demonstrates the effectiveness of the OCR system in overcoming challenges such as varying receipt print quality. This research also offers practical solutions in the form of OCR-based applications that can be used in business environments to improve the efficiency and accuracy of financial transaction data management.*

Keywords : *Information Extraction, Optical Character Recognition (OCR), Digital Image Processing, Image Processing, Financial Transaction Receipt.*

Abstrak : Penelitian ini bertujuan untuk mengotomatiskan proses ekstraksi informasi dari struk transaksi keuangan menggunakan algoritma *Optical Character Recognition* (OCR). Algoritma OCR digunakan untuk mengenali karakter dari citra digital struk transaksi sehingga informasi seperti tanggal, jumlah transaksi, dan rincian lainnya dapat diidentifikasi secara akurat. Dengan menerapkan metode pengolahan citra, penelitian ini berhasil menunjukkan efektivitas sistem OCR dalam mengatasi tantangan seperti kualitas cetakan struk yang beragam. Penelitian ini juga menawarkan solusi praktis berupa aplikasi berbasis OCR yang dapat digunakan dalam lingkungan bisnis untuk meningkatkan efisiensi dan akurasi pengelolaan data transaksi keuangan.

Kata Kunci: Ekstraksi Informasi, *Optical Character Recognition* (OCR), Pengolahan Citra Digital, Pengolahan Gambar, Struk Transaksi Keuangan.

1. PENDAHULUAN

Dengan perkembangan teknologi di Indonesia yang semakin pesat, sistem informasi menjadi sangat penting bagi dunia kerja. Perkembangan yang pesat ini telah memaksa hampir semua bidang aplikasi bisnis untuk menggunakan dan mengembangkan sistem informasi dengan cara yang memungkinkan mereka untuk mengembangkan dan memajukan usaha mereka dengan sangat baik. Perkembangan teknologi yang semakin meningkat ini mempunyai dampak yang sangat besar terhadap perekonomian [1].

Transaksi terjadi di setiap bisnis, baik kecil maupun besar. Jika ada perubahan pada aspek keuangan perusahaan, seperti penurunan atau peningkatan nilai, suatu peristiwa disebut sebagai transaksi.

Kondisi keuangan sebuah perusahaan atau bisnis sangat dipengaruhi oleh perubahan yang terjadi selama transaksi ini. Beberapa contoh kegiatan transaksi di dalam sebuah bisnis termasuk menjual barang, membeli barang bahkan untuk diproduksi, membayar gaji, dan membayar utang bisnis lainnya.

Pada prakteknya dalam usaha kecil dan menengah sering kali lalai dalam melakukan pembukuan. Penyebab utama dari kelalaian tersebut biasanya adalah karena bukti transaksi yang tidak didapatkan sejak awal atau hilang.

Bukti transaksi adalah bukti tertulis yang mencatat atau merekap semua transaksi yang terjadi pada suatu perusahaan atau bisnis. Ini sangat penting untuk mencegah masalah keuangan di kemudian hari [2].

Untuk membantu tim akuntan perusahaan menyusun laporan keuangan, bukti transaksi yang sudah dicatat sangat membantu. Pencatatan bukti transaksi secara teratur dan teratur memiliki banyak keuntungan. Tidak hanya bukti transaksi dapat membantu mengidentifikasi siapa yang bertanggung jawab atas transaksi tertentu, tetapi bukti transaksi juga dapat membantu mencegah kesalahan dalam penghitungan keuangan perusahaan atau bisnis.

Penelitian ini bertujuan untuk mengotomatisasi proses ekstraksi informasi dari struk transaksi keuangan, yang biasanya dilakukan secara manual. Dengan menggunakan algoritma OCR, diharapkan waktu dan usaha yang dibutuhkan untuk memasukkan data dapat diminimalkan.

Salah satu tujuan utama adalah meningkatkan akurasi dalam mengenali dan mengekstrak informasi penting dari struk, seperti tanggal, jumlah transaksi, dan rincian lainnya. Penelitian ini berfokus pada pengembangan metode yang dapat mengatasi tantangan dalam pengenalan karakter yang sering muncul pada struk dengan kualitas cetakan yang bervariasi.

Selain itu, Penelitian ini bertujuan untuk mengembangkan aplikasi berbasis OCR yang dapat digunakan secara praktis dalam lingkungan bisnis atau keuangan. Sistem ini diharapkan dapat diintegrasikan dengan sistem manajemen data yang ada untuk meningkatkan alur kerja.

2. TUJUAN PUSTAKA

Penelitian Terkait

Pada penelitian sebelumnya yang berjudul “Sistem Manajemen Kartu Nama dengan OCR dan Ekstraksi Informasi Otomatis” algoritma OCR diterapkan menggunakan Tesseract untuk mengenali karakter pada kartu nama. Proses ekstraksi informasi dilakukan melalui beberapa tahap, mulai dari *preprocessing* gambar yang bertujuan untuk menghilangkan noise dan memperbaiki orientasi gambar, hingga meningkatkan kualitas *input* bagi algoritma OCR. Setelah itu, dilakukan pengenalan karakter yang memproses gambar menjadi teks, dan dilanjutkan dengan klasifikasi entitas seperti nama, alamat, nomor telepon, serta email. Klasifikasi ini menggunakan pendekatan Naive Bayes yang dikombinasikan dengan aturan berbasis *regular expression*. Penelitian ini menunjukkan hasil yang cukup baik, dengan akurasi pengenalan karakter mencapai 85,1% dan akurasi klasifikasi entitas sebesar 86% [3].

Pada penelitian yang berjudul "Aplikasi Ekstraksi Data Kartu Vaksin Berbasis Web Menggunakan Metode OCR", algoritma OCR digunakan untuk mengekstraksi informasi dari gambar kartu vaksin dan menyajikan hasilnya dalam format teks pada sebuah *website*. Penelitian ini menggunakan teknologi *serverless* yang memungkinkan pengolahan data tanpa memerlukan infrastruktur server fisik, dengan memanfaatkan layanan seperti *Google Cloud Storage* dan *Google Cloud Functions*. Penelitian ini menguji sistemnya menggunakan teknik pengujian kotak hitam dan putih untuk memastikan setiap komponen berjalan sesuai dengan fungsinya. Menurut hasil pengujian, sistem dapat menampilkan hasil ekstraksi dengan baik, selama kualitas gambar yang diunggah memenuhi syarat, seperti pencahayaan yang cukup dan tanpa blur [4].

Pengolahan Citra Digital

Pengolahan gambar digital membahas peningkatan kualitas gambar (perbaikan kontras, peningkatan warna, resolusi gambar), transformasi gambar (translasi, rotasi transformasi, skala, geometrik), dan pemilihan citra ciri (ciri gambar) yang paling cocok untuk analisis. Ini juga mencakup penyimpanan dan pengiriman data yang telah dikompresi dan direduksi sebelumnya, serta waktu proses data. Di bawah ini adalah diagram sederhana dari proses pengolahan [5].

Citra digital adalah representasi objek yang diolah di dalam perangkat komputer dari berbagai perangkat digital. Data pada citra digital berisi berbagai informasi, sehingga digunakan sebagai barang bukti dalam konferensi dan domain digital forensik [6].

Transformasi Warna

Dalam pengolahan gambar digital, warna RGB adalah warna yang paling umum dan biasa digunakan oleh semua benda sehari-hari. Terdapat tiga warna dalam model warna RGB: merah (*red*), hijau (*green*), dan biru (*blue*). Dapat membuat berbagai warna dengan menggabungkan warna-warna ini dengan berbagai cara. Model warna RGB didasarkan pada ide penambahan cahaya yang kuat. Dasar merah, hijau, dan biru. Jika ada ruangan yang sama sekali tanpa sinar, tempat itu gelap sepenuhnya. Mata kita tidak menyerap gelombang cahaya, atau RGB (0, 0, 0). Namun, jika kita menambahkan cahaya merah, misalnya RGB (255, 0, 0), Segala sesuatu di dalam ruangan akan berwarna merah. Jadi jika kita mengubah cahaya menjadi hijau atau biru. Seperti yang diketahui semua orang, sistem warna RGB, juga dikenal sebagai *Red*, *Green*, dan *Blue*, digunakan secara luas untuk pewarnaan tampilan digital dan banyak digunakan untuk monitor, video, layar telepon, dll. Sistem warna RGB terdiri dari seratus persen merah, seratus persen hijau, dan seratus persen biru, yang menghasilkan 100% putih RGB tidak mengandung hitam [7].

OCR (Optical Character Recognition)

OCR (*Optical Character Recognition*) adalah teknologi/algoritma pemindaian yang memanfaatkan gambar untuk mendeteksi teks, angka, dan pola karakter [8]. Ia dapat membedakan gambar yang ditransmisikan dari rekaman, yang dapat berupa tulisan tangan atau digital dan dapat diproses. Sistem komputer yang dikenal sebagai OCR (*Optical Character Recognition*) dapat membaca karakter dari printer (*printer* atau mesin tik) atau tulisan tangan. Aplikasi OCR biasanya digunakan untuk mengidentifikasi gambar teks dan mengonversinya untuk menyimpan *file* [9].

Python

Bahasa pemrograman python populer dalam bidang data karena analisis mudah dipelajari dan digunakan oleh orang-orang dari semua usia. Selain itu, bahasa ini memiliki banyak perpustakaan yang berguna yang dapat digunakan oleh siapa saja di berbagai sistem operasi, atau dengan kata lain bersifat *open source*. Beberapa perpustakaan python adalah NumPy, Pandas, dan Matplotlib. Selain itu, python sangat mudah diintegrasikan dengan berbagai teknologi yang berkaitan dengan analisis data, seperti *framework web*, *database*, dan alat besar data, yang memungkinkan pengaksesan dan pengelolaan data dari berbagai sumber [10].

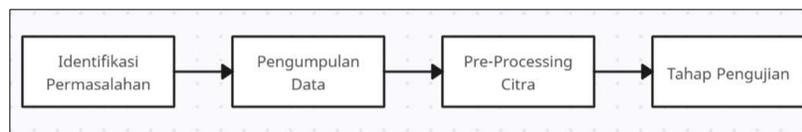
3. METODE PENELITIAN

Analisis

Berdasarkan permasalahan yang muncul penelitian ini terfokus pada tiga aspek utama. Pertama, apakah metode OCR mampu mengekstrak teks berupa informasi dari struk transaksi keuangan. Selanjutnya, sejauh mana hasil dan ketepatan akurasi yang dihasilkan oleh program dengan menggunakan metode OCR dapat memberikan informasi yang tepat terkait struk transaksi keuangan.

Tahap Pelaksanaan

Tahap pelaksanaan penelitian dengan menggunakan metode pengolahan citra. Tahapan dalam metode pelaksanaan ini ditunjukkan seperti gambar berikut:



Gambar 1. Tahap pelaksanaan penelitian

Identifikasi Permasalahan

Pada tahap ini, peneliti akan melakukan perencanaan terhadap kebutuhan yang diperlukan untuk mengekstrak informasi dari struk transaksi keuangan menggunakan pengolahan citra. Peneliti akan turun langsung ke lokasi untuk mengamati bagaimana proses penginputan data dari struk dilakukan secara manual, di mana setiap transaksi masih diinput satu per satu dengan cara diketik. Setelah mengidentifikasi permasalahan dalam proses input data yang memakan waktu dan rentan terhadap kesalahan, peneliti akan menentukan metode yang tepat untuk memecahkan masalah tersebut dengan menerapkan algoritma OCR. Dengan demikian, penelitian ini bertujuan untuk meningkatkan efisiensi dan akurasi dalam pengolahan data transaksi keuangan.

Pengumpulan Data

Pada tahap ini, Peneliti mulai mendapatkan informasi yang diperlukan, seperti mengumpulkan foto struk transaksi keuangan. Struk tersebut difoto menggunakan kamera ponsel melalui aplikasi *CamScanner* dengan pencahayaan yang memadai dan tanpa kabur, sehingga memudahkan proses deteksi teks menggunakan metode OCR. Peneliti memastikan bahwa kualitas gambar yang dihasilkan cukup baik untuk meningkatkan akurasi ekstraksi informasi dari struk transaksi.

Pre-Processing Citra

Pada tahap ini, peneliti melakukan tahap pertama, yaitu melakukan cropping pada foto struk transaksi keuangan. Foto struk tersebut dipotong agar fokus deteksi hanya pada bagian informasi penting, sementara elemen lain seperti barcode dihapus untuk memudahkan proses ekstraksi teks. Setelah proses *cropping* dan penghilangan elemen yang tidak diperlukan selesai, citra struk transaksi diolah menggunakan teknik *Optical Character Recognition* (OCR). Bagian teks yang terdeteksi akan diberi kotak merah untuk menandai hasil ekstraksi. Setelah teks berhasil diekstrak, hasilnya akan otomatis tersimpan dalam format *Microsoft Word* untuk memudahkan akses dan pengolahan lebih lanjut.

Tahap Pengujian

Pada tahap ini, dilakukan pengujian terhadap hasil pengenalan karakter pada struk transaksi keuangan menggunakan metode *Optical Character Recognition* (OCR). Tujuannya adalah untuk mengetahui apakah metode OCR cocok dalam pengenalan karakter tekstual untuk membaca informasi dari struk transaksi keuangan. Pengujian ini akan memberikan data tentang seberapa akurat metode OCR dalam membaca informasi dari struk transaksi keuangan, hasilnya kemudian dicatat dalam bentuk tabel hasil pengujian.

4. HASIL DAN PEMBAHASAN

Hasil

Adapun hasil penelitian dari ekstraksi teks menggunakan metode *Optical Character Recognition* (OCR) pada struk transaksi keuangan adalah sebagai berikut:

- a. Dapat mengekstrak informasi penting dari struk transaksi keuangan melalui pengolahan citra dengan membaca citra digital.
- b. Data yang digunakan dalam aplikasi ini adalah foto digital struk transaksi, di mana pada penelitian ini foto digital yang digunakan sebanyak 20 foto dengan pencahayaan yang cukup.
- c. Dapat menerapkan metode OCR untuk membaca kembali informasi yang terdapat pada struk transaksi keuangan dengan tingkat akurasi yang tinggi.

Pembahasan

Proses pelaksanaan penelitian untuk mengekstrak informasi dari gambar menjadi teks ini menggunakan *Visual Studio Code* yang terinstal pada sistem operasi windows 11. Representasi warna citra yang digunakan adalah citra warna asli (RGB), yang memungkinkan analisis yang lebih akurat terhadap informasi yang terkandung dalam struk transaksi keuangan.

a. Proses Pengambilan Gambar

Langkah awal yang dilakukan peneliti adalah mengambil citra struk transaksi keuangan menggunakan kamera ponsel dengan pencahayaan yang memadai. Foto struk diambil menggunakan aplikasi tambahan, yaitu *CamScanner*, untuk memastikan hasil pemindaian tidak blur dan terlihat jelas agar mudah diproses nantinya. Setelah selesai difoto, 20 struk transaksi ini disimpan dalam satu folder yang sama untuk memudahkan pengolahan lebih lanjut. Adapun contoh struk transaksi keuangan sebagai berikut digunakan dalam penelitian ini:

SURYA ELEKTRONIK		
Jl. Veteran Utara No.248B		
Telp. 0411 3628529		
Sabtu, 16 November 2019		
13:56:30		
<hr/>		
PPG		
500.000	X1	500.000
Pulse Sensor		
150.000	X1	150.000
Pointer		
225.000	X1	225.000
<hr/>		
Total Belanja	Rp.	875.000
Tunai	RP.	900.000
Kembali	Rp.	25.000
Terima Kasih		

Gambar 2. Citra struk

Setelah mendapatkan objek yang akan diteliti, hasil citra berupa struk transaksi keuangan ini diedit secara manual dengan melakukan proses cropping atau pemotongan gambar. Proses ini bertujuan untuk mengambil hanya bagian informasi penting dari struk, serta menghapus elemen seperti barcode untuk memudahkan dalam mengekstrak informasi yang terdapat pada struk transaksi. Adapun contoh citra struk transaksi keuangan yang digunakan dalam penelitian ini adalah sebagai berikut:

SURYA ELEKTRONIK		
Jl. Veteran Utara No.248B		
Telp. 0411 3628529		
Sabtu, 16 November 2019		
13:56:30		
PPG		
500.000	X1	500.000
Pulse Sensor		
150.000	X1	150.000
Pointer		
225.000	X1	225.000
<hr/>		
Total Belanja	Rp.	875.000
Tunai	RP.	900.000
Kembali	Rp.	25.000
Terima Kasih		

Gambar 3. Objek struk

b. Langkah pengolahan citra

Setelah kartu tanda mahasiswa melalui tahap pemotongan dan penghapusan, lanjut ke tahap pemrosesan citra. Citra yang bertipe *jpeg* dengan intensitas warna RGB terletak dalam satu folder bernama *cropped*. Lalu, semua citra tersebut diproses secara bersamaan untuk mempersiapkan mereka untuk analisis lebih lanjut menggunakan metode OCR.

c. Proses Pembacaan Informasi pada Struk Transaksi Keuangan

Setelah foto diinput, langkah selanjutnya adalah melakukan ekstraksi teks yang diambil dari gambar melalui penggunaan teknik *Optical Character Recognition* (OCR). Proses kerja OCR ini dilakukan secara otomatis melalui berbagai langkah, seperti pra-pemrosesan, segmentasi, ekstraksi fitur, dan pengenalan teks dari gambar. Adapun hasil dari ekstraksi citra menggunakan algoritma OCR yaitu:

```
Teks yang Diekstraksi:  
SURYA ELEKTRONIK  
Jl. Veteran Utara No.248B  
Telp. 0411 3628529  
Sabtu, 16 November 2019  
  
13:56:30  
PPG  
500.000 X21 500.000  
Pulse Sensor  
150.000 x1 150.000  
Pointer  
Pointer  
225.000 X21 225.000  
  
Total Belanja Rp. 875.000  
Tunai RP. 900.000  
Kembali Rp. 25.000  
  
Terima Kasih
```

Gambar 4. Hasil ekstraksi teks

d. Pengujian Sistem

Dalam penelitian ini terdapat 20 citra uji struk transaksi keuangan yang digunakan untuk mengetahui tingkat keberhasilan pengolahan citra. Untuk itu, akan digunakan perhitungan Persentase Keberhasilan (*Success Rate Percentage*) untuk menentukan seberapa akurat hasil ekstraksi teks dari gambar yang diperoleh melalui metode *Optical Character Recognition* (OCR).

5. KESIMPULAN DAN SARAN

Kesimpulan

- a. Algoritma OCR mampu mengekstrak informasi penting seperti tanggal, jumlah transaksi, dan rincian lainnya dari struk transaksi keuangan dengan tingkat akurasi yang tinggi.
- b. Metode pengolahan citra yang digunakan, seperti *cropping* dan transformasi warna, efektif dalam meningkatkan akurasi ekstraksi.
- c. Aplikasi berbasis OCR yang dikembangkan mampu menghemat waktu dan meminimalkan kesalahan penginputan data secara manual.

Saran

- 1) Penelitian ini dapat ditingkatkan dengan mengintegrasikan algoritma *machine learning* untuk meningkatkan akurasi pengenalan karakter pada struk dengan kualitas rendah.
- 2) Pengujian pada skala data yang lebih besar diperlukan untuk mengukur keandalan sistem dalam berbagai kondisi nyata.

DAFTAR PUSTAKA

- Alfatul Hisabi, A., Azura, A., Lutfiah, D., & Nurbaiti. (2022). Perkembangan sistem informasi manajemen (SIM) di Indonesia. *Juremi: Jurnal Riset Ekonomi*, 1(4), 364–371. <https://doi.org/10.53625/juremi.v1i4.775>
- Darmawan, R., Nasuha, A., Zaman, L., & Armanto, H. (2021). Sistem manajemen kartu nama dengan OCR dan ekstraksi informasi otomatis. *Journal of Intelligent Systems and Computing*, 3(2), 61–72. <https://doi.org/10.52985/insyst.v3i2.194>
- Fareza, A. D., Toyib, R., Studi, P., Teknik Informatika, F., & Universitas M. Bengkulu. (2024). Analisis dan implementasi citra digital pada buku Ishihara untuk melihat tingkat buta warna. *Jurnal Teknologi*, 8(5), 11043–11049.
- Khairunnisak, K., Ashari, H., & Kuncoro, A. P. (2020). Analisis forensik untuk mendeteksi keaslian citra digital menggunakan metode NIST. *Jurnal Resist (Rekayasa Sistem Komputer)*, 3(2), 72–81. <https://doi.org/10.31598/jurnalresistor.v3i2.634>
- Nafsin, M., Qashlim, A., & Khairat, U. (2022). Sistem informasi data siswa berbasis OCR (Optical Character Recognition) pada SMK Bina Harapan. *Jurnal Pengajaran Konferensi Ser*, 4(1), 412. <https://doi.org/10.35329/jp.v4i1.2201>
- O. C. R. Untuk, & E. Informasi. (2024). Penerapan algoritma OCR untuk ekstraksi informasi dari citra kartu tanda mahasiswa (KTM). *Jurnal Ilmiah Teknologi*, 8(5), 11004–11009.
- Pradila, R., Putri, P., Latifah, N., & S. W. U. (2024). Pengelolaan bukti transaksi pengeluaran berbasis web di PT Tiga Sahabat Teknik. *Jurnal Ilmu Teknologi*, 1(1), 36–51.
- Sambi, A. M. T. I., & Ua, S. (2023). Penggunaan bahasa pemrograman Python dalam analisis faktor penyebab kanker paru-paru. *Jurnal Publikasi Teknologi Informasi*, 2(2), 88–99. <https://doi.org/10.55606/juipi.v2i2.1742>
- Santi, S., Susanto, C., Muhardi, M., Patasik, M., & Nurlina, N. (2024). Penerapan algoritma K-Nearest Neighbor (KNN) dalam pengklasifikasian tingkat kematangan buah nangka berdasarkan citra warna kulit. *Digital Transformation & Technology*, 4(1), 685–692. <https://doi.org/10.47709/digitech.v4i1.4550>
- Wahyuddin, W., & Hasim, A. (2023). Aplikasi ekstraksi data kartu vaksin berbasis web menggunakan metode OCR. *Jurnal Sintaks Logika*, 3(2), 53–57. <https://doi.org/10.31850/jsilog.v3i2.2525>