

PERANCANGAN METODE OCR (*OPTICAL CHARACTER RECOGNITION*) PADA APLIKASI PEMANTAUAN PENGUNAAN BBM MENGUNAKAN *TESSERACT* (STUDI KASUS DI PT.ABACUS)

Ardi Pamungkas R

Program Studi Teknik Komputer, Politeknik Sukabumi

Jl. Babakan Sirna No.25, Benteng, Kec. Warudoyong, Kota Sukabumi, Jawa Barat 43132

ardipamungkas655@gmail.com

Abstrak

Pengelolaan dan pemantauan penggunaan Bahan Bakar Minyak (BBM) merupakan aktivitas penting dalam operasional perusahaan untuk memastikan efisiensi biaya dan transparansi laporan, namun proses pencatatan manual dari struk pembelian BBM sering menyebabkan kesalahan input dan keterlambatan pelaporan. Penelitian ini bertujuan mengimplementasikan metode *Optical Character Recognition (OCR)* menggunakan *Tesseract* pada aplikasi pemantauan penggunaan BBM guna mengotomatisasi proses ekstraksi data dari struk BBM, dengan studi kasus di **PT. Abacus**. Metode *OCR Tesseract* diintegrasikan ke dalam aplikasi untuk membaca informasi penting seperti tanggal pengisian, liter, harga per liter, dan total biaya dari struk BBM yang diunggah oleh pengguna. Hasil pengujian menunjukkan bahwa *OCR* mampu mengekstraksi data dengan tingkat akurasi yang baik terutama pada struk dengan kualitas gambar dan pencahayaan memadai. Implementasi ini membantu mempercepat proses input data, mengurangi kesalahan manual, serta meningkatkan efisiensi monitoring pengeluaran BBM perusahaan. Dengan demikian, penerapan *OCR Tesseract* memberikan solusi otomatisasi yang efektif dalam proses administrasi dan pengawasan penggunaan BBM di PT. Abacus..

Kata Kunci: *OCR, Tesseract, Pemantauan BBM, Ekstraksi Data, Struk BBM*

Abstract

Managing and monitoring fuel usage is a crucial activity in company operations to ensure cost efficiency and reporting transparency. However, the manual recording of fuel purchase receipts often leads to input errors and reporting delays. This study aims to implement the Optical Character Recognition (OCR) method using Tesseract in a fuel usage monitoring application to automate the data extraction process from fuel receipts, using a case study at PT. Abacus. The Tesseract OCR method was integrated into the application to read important information such as fill date, liters, price

per liter, and total cost from user-uploaded fuel receipts. Test results showed that OCR was able to extract data with a high level of accuracy, especially on receipts with adequate image quality and lighting. This implementation helped speed up the data input process, reduce manual errors, and improve the efficiency of monitoring the company's fuel expenditure. Thus, the implementation of Tesseract OCR provides an effective automation solution for the administration and monitoring of fuel usage at PT. Abacus.

Keywords: OCR, Tesseract, Fuel Monitoring, Data Extraction, Fuel Receipts

I. PENDAHULUAN

Dengan perkembangan teknologi di Indonesia yang semakin pesat, sistem Informasi menjadi sangat penting bagi dunia kerja. [1]. 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 perusahaan mereka dengan sangat baik. Perkembangan teknologi yang semakin meningkat ini mempunyai dampak yang sangat besar terhadap berbagai macam bidang. Ada dalam bidang pendidikan, ke pemerintahan, kesehatan, hiburan, dan bahkan dalam bidang bisnis. [2]. Salah satu perkembangan teknologi di bidang bisnis adalah monitoring data Bahan bakar minyak (BBM) [3].

Berdasarkan hasil wawancara, di PT. Abacus khususnya dibagian operasional kendaraan dalam hal menggunakan pengolahan data masih menggunakan sistem yang bersifat manual, dimana proses pengolahan data pengguna BBM menggunakan sistem pencatatan langsung yang berupa dokumen (arsip) misalnya nota pembelian Bahan bakar minyak (BBM). [4]. Saat ini sistem yang digunakan dalam penyajian data ataupun penginputan data masih menggunakan cara konvensional. Cara konvensional yang kami maksud yaitu penginputan yang dilakukan menggunakan sebuah aplikasi *Google Form* yang dilakukan pengetikan sebagai proses pencatatan laporan,

dan cara itu sangat tidak efektif dan efisien [5]. Dengan membangun sebuah sistem Informasi berbasis *web* yang terintegrasi dan dapat diakses secara *realtime*. Sistem ini dirancang untuk menggantikan proses pencatatan manual yang selama ini dilakukan melalui dokumen fisik atau Formulir digital sederhana seperti *Google Form*, yang kurang efektif dan efisien dalam jangka panjang. [6]. Sistem informasi yang dikembangkan akan menggunakan *Framework Laravel* sebagai *Backend* dan *MySQL* sebagai basis data utama.

Beberapa fitur utama yang akan disediakan antara lain adalah *Form* untuk mencatat penggunaan BBM dengan cara unggah bukti pembelian dalam bentuk foto, manajemen data pengguna dan kendaraan lalu unggahan tersebut akan di konversikan menjadi teks menggunakan metode *OCR*, serta *dashboard* untuk memantau dan menganalisis data BBM berdasarkan pengguna, kendaraan, dan periode waktu tertentu [7].

Beberapa penelitian yang telah dilakukan terkait dengan metode *Optical Character Recognition (OCR)* dengan tujuan yang berbeda-beda. Latifa Khoirani, Rino Ariansyah, Supiyandi Supiyandi (2024) Melakukan penelitian yang membuat aplikasi untuk mengotomatisasi proses ekstraksi Informasi dari struk transaksi keuangan, yang biasanya dilakukan secara manual. [9]. Dengan menggunakan algoritma *Optical Character Recognition (OCR)*, diharapkan waktu dan usaha yang dibutuhkan untuk memasukkan data dapat

diminimalkan [8].

Septio P, Yulianto S, Prasetyo J (2023) Melakukan penelitian yang membuat aplikasi untuk mengatasi masalah menginput data digital dengan jumlah banyak pada Aplikasi Validasi Dokumen Tagihan Pembelian Barang Secara Digital, dan pada penelitian ini menggunakan metode *Optical Character Recognition (OCR) Tesseract* [9].

Perbedaan penelitian sebelumnya dengan tugas akhir ini terletak pada penggunaannya dan juga implementasinya yaitu penggunaan metode *Optical Character Recognition (OCR)* yang akan melakukan *scanning* pada Struk BBM dan lalu akan terintegrasi kepada Aplikasi *Web* berbasis *Framework Laravel* yang memiliki kemampuan untuk memudahkan manajemen pendataan Struk Bahan bakar minyak (BBM).

Tugas akhir ini bertujuan untuk merancang aplikasi pemantauan penggunaan BBM berbasis android yang dilengkapi dengan fitur pemindaian otomatis struk BBM menggunakan metode *Optical Character Recognition (OCR)* dengan *Tesseract*. Melalui penerapan teknologi *OCR*, sistem ini diharapkan mampu mengekstraksi informasi penting pada struk BBM secara cepat dan akurat sehingga proses input data tidak lagi dilakukan secara manual. Dengan adanya sistem ini, proses monitoring pengeluaran BBM dapat dilakukan dengan lebih efisien, data yang diterima menjadi lebih valid, serta meminimalkan kesalahan pencatatan. Implementasi sistem ini tidak hanya

meningkatkan kualitas pengelolaan administrasi penggunaan BBM di PT. Abacus, tetapi juga mendukung transparansi, akurasi, dan efektivitas dalam proses operasional perusahaan.

II. METODE PENELITIAN

A. *Optical Character Recognition (OCR)*

Optical Character Recognition (OCR) adalah salah satu teknik pada bidang pengolahan citra dan *computer vision* yang bertujuan untuk mengidentifikasi citra huruf atau pun angka yang dikonversi menjadi bentuk *File* tulisan. Sistem pengenalan huruf ini dapat meningkatkan fleksibilitas atau kemampuan dan kecerdasan sistem komputer. Sistem pengenalan huruf yang cerdas sangat membantu usaha besarbesaran yang saat ini dilakukan banyak pihak yakni usaha digitalisasi Informasi dan pengetahuan, misalnya dalam pembuatan koleksi pustaka digital, koleksi sastra kuno digital, dan lain-lain. Secara umum, proses ini menggunakan beberapa tahap yaitu :

A. *File input*

Memasukkan *File* yang akan diolah dengan *Format *.bmp* atau **.jpg*

B. *Preprocessing*

Preprocessing adalah proses yang bertujuan untuk menghilangkan bagian-bagian dari suatu citra yang tidak diperlukan pada gambar input pada proses selanjutnya.

C. *Segmentation*

Segmentation merupakan suatu proses untuk memisahkan suatu area pengamatan (*Region*) pada setiap karakter yang dideteksi.

D. Normalization

Normalization merupakan suatu proses untuk merubah suatu dimensi *Region* dari setiap karakter dan ketebalan karakter.

E. Feature extraction

Feature extraction adalah suatu proses untuk mengambil suatu ciri tertentu dari karakter yang diamati.

F. Recognition

Recognition merupakan suatu proses untuk mengenali karakter yang diamati dengan cara membandingkan ciri-ciri dari suatu karakter yang diperoleh dengan ciri-ciri karakter yang ada pada basis data yang ada[5]

B. Tesseract Optical Character Recognition (OCR)

Tesseract OCR adalah sebuah library open-source yang banyak digunakan dalam proses pengenalan karakter optik (Optical Character Recognition) untuk mengubah citra biner—baik berupa gambar hasil pemindaian, foto struk, dokumen, maupun teks tercetak—menjadi teks digital yang dapat diproses lebih lanjut. *Tesseract* awalnya dikembangkan oleh Hewlett-Packard (HP), kemudian dilanjutkan dan dipelihara oleh Google sehingga terus mengalami pengembangan signifikan, baik dari segi akurasi pengenalan karakter maupun fleksibilitas penggunaannya dalam berbagai aplikasi modern.

Sebagai sebuah library, *Tesseract* dapat

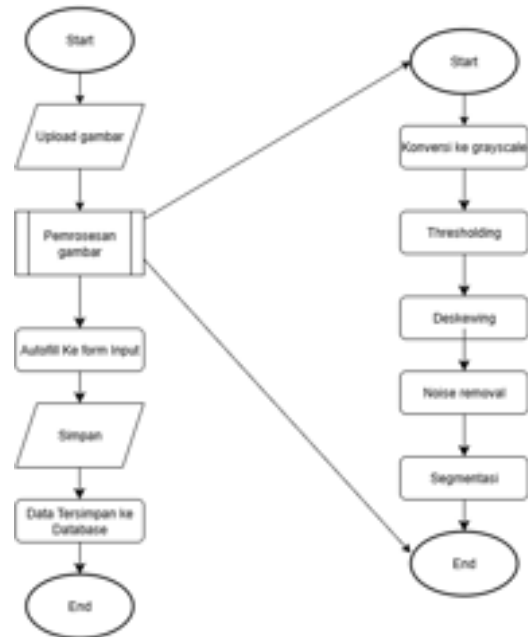
dioperasikan secara langsung melalui perintah baris (command-line), atau digunakan oleh pengembang melalui API untuk mengintegrasikan fitur *OCR* ke dalam aplikasi berbasis web, desktop, maupun mobile. Meskipun *Tesseract* tidak menyediakan antarmuka grafis bawaan (GUI), tersedia banyak GUI pihak ketiga yang mempermudah pengguna dalam memproses dokumen tanpa perlu melakukan konfigurasi manual. Fleksibilitas *Tesseract* juga didukung oleh keberadaannya di berbagai bahasa pemrograman dan kerangka kerja melalui wrapper seperti untuk Python, Java, C#, PHP, dan Node.js, sehingga menjadikannya salah satu solusi *OCR* yang paling banyak diadopsi secara global.

Tesseract mampu membaca dokumen dalam berbagai bahasa, termasuk bahasa Indonesia, karena memiliki model bahasa (traineddata) yang dapat diunduh dan digunakan sesuai kebutuhan. Library ini dapat bekerja dalam dua pendekatan utama: pertama, memanfaatkan sistem analisis tata letak internal untuk mengenali teks pada dokumen besar; kedua, bekerja bersama detektor teks eksternal seperti OpenCV atau EAST Text Detector untuk mengidentifikasi lokasi teks pada gambar satu baris atau fragmen tertentu sebelum dilakukan proses *OCR*. Kemampuan ini membuat *Tesseract* cukup fleksibel untuk digunakan dalam berbagai skenario, mulai dari scanning dokumen resmi hingga ekstraksi informasi dari struk belanja atau plat nomor kendaraan.

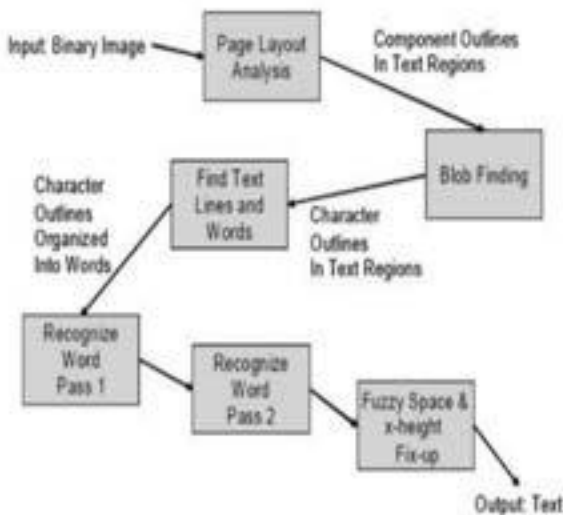
Secara teknis, proses pembacaan karakter oleh *Tesseract OCR* melalui beberapa tahapan,

seperti pra-pemrosesan citra, segmentasi, normalisasi, hingga klasifikasi karakter. *Tesseract* menggunakan metode algoritma matrix matching dan teknik feature extraction untuk mengenali pola setiap karakter dalam citra. Matrix matching memungkinkan sistem membandingkan pola matriks piksel dengan data pelatihan yang ada, sedangkan ekstraksi fitur membantu mengidentifikasi elemen penting dari karakter seperti garis, lengkungan, hingga arah goresan. Kombinasi kedua metode ini memungkinkan *Tesseract* menghasilkan akurasi pengenalan teks yang lebih tinggi, terutama jika citra yang diproses memiliki kualitas yang baik.

flowchart untuk memetakan alur proses secara visual. Berikut adalah gambar 2. yang menunjukkan proses alur dari *OCR* tersebut :



Gambar 2. Flowchart Proses Aplikasi dan Pemrosesan gambar OCR



Gambar 1. Arsitektur Mesin Tesseract [10]

C. Langkah dan prosedur penerapan metode OCR

Dalam pengembangan aplikasi ini , langkah awal melibatkan penggunaan diagram

Flowchart ini menggambarkan langkah-langkah dari pemilihan Gambar, Pemrosesan Gambar menggunakan Metode *Tesseract OCR*, dari konversi ke grayscale hingga Penyimpanan data ke database. Selain itu ada penjelasan deskriptif tentang perancangan dan peran metode *OCR* di dalam aplikasi ini sebagai berikut :

1. Input Gambar

Pada tahap ini pengguna memberikan gambar struk BBM. Gambar ini bisa berasal dari:

- a. Kamera (langsung diambil)
- b. Galeri (diambil dari penyimpanan perangkat)

Gambar tersebut biasanya berupa foto

struk pembelian BBM yang mengandung teks seperti total harga, volume liter, dan Informasi lainnya.

2. Pra-pemrosesan Gambar (*Preprocessing*)

Agar *Tesseract* bisa mengenali teks dengan akurat, gambar harus melalui proses persiapan, seperti:

- a. Konversi ke *Grayscale* → Mengubah gambar berwarna ke hitam-putih untuk fokus pada bentuk huruf. Rumus untuk *Grayscale* (Secara umum) :

$$Y = 0.299R + 0.587G + 0.114B$$

- b. *Thresholding* → Menghilangkan latar belakang dan hanya menyisakan teks yang kontras. Contoh :

Jika pixel > 127 → putih (255)

Jika pixel ≤ 127 → hitam (0)

- c. *Deskewing* → Meluruskan gambar jika miring agar huruf tidak terdistorsi. *Deskewing* ini menggunakan deteksi sudut kemiringan (*angle*) dari kontur teks, lalu melakukan rotasi balik.

Rumus Transformasi Rotasi :

$$x' = x \cdot \cos\theta - y \cdot \sin\theta$$

$$y' = x \cdot \sin\theta + y \cdot \cos\theta$$

- d. *Noise Removal* → Menghapus titik-titik atau garis yang mengganggu agar teks lebih jelas.

3. Segmentasi

Tesseract memecah gambar menjadi bagian-bagian kecil untuk dianalisis. Proses ini meliputi:

- a. Deteksi baris teks
- b. Pemisahan kata dan huruf
- c. Identifikasi blok teks atau kolom

Sistem mencoba mengenali struktur *layout* dari teks seperti: kalimat, spasi, paragraf, dan lainnya.

4. Ekstraksi fitur

Setelah teks di-segmentasi, *Tesseract* akan menganalisis setiap karakter atau huruf. Ia membandingkan bentuk karakter dari gambar dengan bentuk karakter yang disimpan dalam model pelatihan huruf (*alphabet recognition model*). Contohnya, bentuk seperti “1” atau “l” akan dicek berdasarkan kontur dan pola pixel-nya.

5. Pengenalan Teks (*Recognition*)

Dengan model pembelajarannya, *Tesseract* mencocokkan bentuk huruf dari gambar dengan huruf-huruf standar yang dikenalnya (misalnya huruf A-Z, angka 0-9, dan simbol).

6. Hasil teks dikirim ke aplikasi

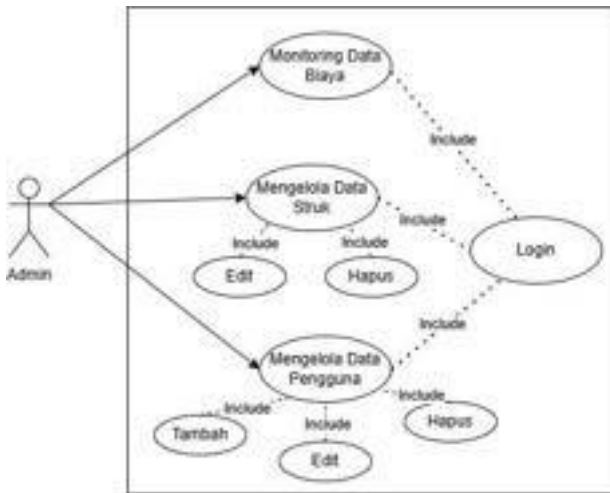
Setelah teks dikenali, *Tesseract* akan mengirimkannya sebagai hasil (output).

Aplikasi kemudian dapat:

1. Menampilkan teks tersebut ke pengguna.
2. Menemukan bagian penting seperti “Rp” dan “Liter” dengan algoritma tambahan seperti pencarian kata kunci.
3. Mengisi *Form* secara otomatis dengan nilai-nilai tersebut.

F. Perancangan Perangkat Lunak

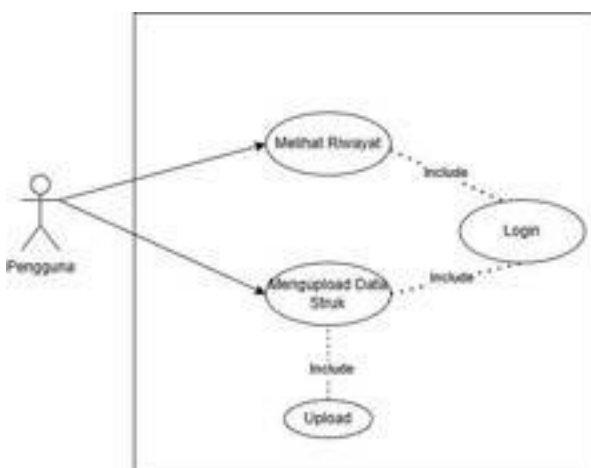
1. Use case Diagram



Gambar 3. Use Case Admin

Use case diagram pada gambar 3. menjelaskan use case admin sebagai berikut:

1. Admin harus login terlebih dahulu.
2. Admin dapat mengelola data Struk
3. Admin dapat mengelola data pengguna/Pengguna
4. Admin Monitoring Data biaya
5. Melihat Mengelola Data Riwayat



Gambar 4. Use Case User

Adapun penjelasan use case diagram pengguna yang di tunjukan pada gambar 4. adalah sebagai berikut:

1. Pengguna harus login terlebih dulu.
2. Pengguna dapat melihat riwayat.
3. Pengguna dapat melihat status upload.
4. Pengguna diharuskan untuk mengupload struk
5. Pengguna akan melihat status di menu status

III. HASIL DAN PEMBAHASAN

A. Perancangan Arsitektur Sistem

Perancangan sistem pemantauan penggunaan BBM dilakukan dengan membangun dua platform terintegrasi, yaitu aplikasi Android berbasis Expo untuk pengguna lapangan (deliveryman/driver) dan aplikasi web untuk admin perusahaan. Arsitektur sistem dirancang menggunakan konsep *client-server*, di mana aplikasi Android berfungsi sebagai pengumpul data struk BBM yang diunggah oleh pengguna, sementara server memproses data menggunakan metode *OCR Tesseract* dan menyimpannya ke dalam basis data. Aplikasi web menyediakan fitur pemantauan, validasi data, serta visualisasi pengeluaran BBM secara real-time untuk admin PT. Abacus. Dengan arsitektur ini, alur kerja menjadi efisien dan dapat diakses dari berbagai perangkat.

B. Perancangan Proses OCR Menggunakan Tesseract

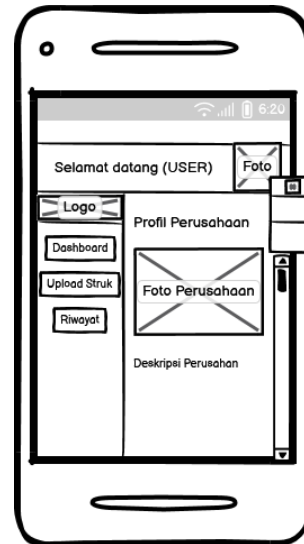
Proses *OCR* dirancang untuk berjalan di sisi server setelah pengguna Android mengunggah foto struk. Tahapan *pipeline OCR* meliputi:

1. Pengambilan foto struk melalui aplikasi Android Expo,
2. Upload otomatis ke server,
3. Pre-processing citra, seperti *resize*, *grayscale*, *thresholding*, dan peningkatan kontras,
4. Ekstraksi teks menggunakan *Tesseract*,
5. Parsing informasi yang mencakup tanggal, liter, harga, dan total biaya,
6. Penyimpanan hasil ke database,
7. Pengiriman hasil kembali ke aplikasi web untuk ditampilkan kepada admin.

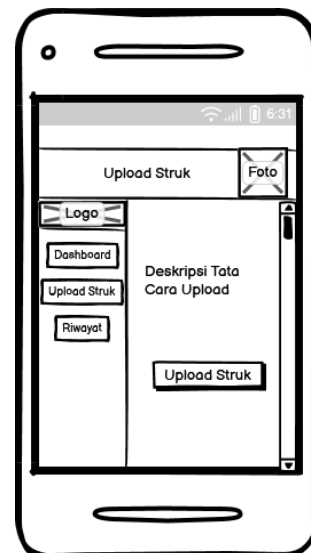
Perancangan ini mempertimbangkan variasi kualitas foto dari kamera Android, sehingga tahapan pra-pemrosesan dibuat fleksibel agar dapat meningkatkan akurasi *OCR* meskipun kualitas gambar tidak ideal.

C. Perancangan Antarmuka Aplikasi

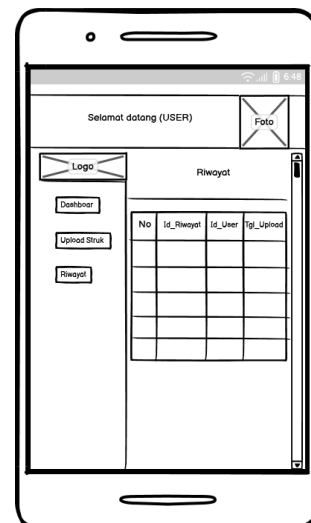
Aplikasi Android dirancang untuk menghasilkan pengalaman penggunaan yang sederhana dan efisien. Fitur utama meliputi:



Gambar 5. Halaman Awal



Gambar 6. Halaman Upload



Gambar 7. Halaman Riwayat

D. Perancangan Basis Data

Basis data dirancang untuk mendukung interaksi antara aplikasi Android dan aplikasi web melalui API. Beberapa tabel utama yang dirancang meliputi:

- tabel *_users* (driver/deliveryman dan admin),
- tabel *_struk* (menyimpan foto struk yang diunggah),
- tabel *_OCR_results* (menyimpan hasil ekstraksi *OCR*),
- tabel *_riwayat* (log aktivitas upload dan verifikasi).

Relasi antar-tabel dibuat untuk mempermudah proses pelacakan dan audit setiap penggunaan BBM oleh karyawan.

IV. KESIMPULAN

Berdasarkan hasil perancangan dan implementasi, aplikasi pemantauan penggunaan BBM berbasis Android dengan integrasi metode *OCR Tesseract* serta aplikasi web untuk admin berhasil memenuhi tujuan penelitian, yaitu menghasilkan sistem yang mampu mengotomatisasi proses pembacaan struk BBM, mempercepat input data, meningkatkan akurasi pencatatan, dan mempermudah proses monitoring pengeluaran BBM di PT. Abacus. Pengguna Android dapat mengunggah foto struk secara langsung, sementara server memproses teks menggunakan *Tesseract* dan menampilkan hasilnya secara real-time pada dashboard web

admin, sehingga alur kerja menjadi lebih efisien dan transparan. Untuk pengembangan selanjutnya, sistem dapat ditingkatkan dengan menambahkan model *OCR* berbasis deep learning agar lebih akurat pada kondisi gambar buruk, integrasi GPS untuk validasi lokasi pengisian BBM, serta pengembangan fitur analitik untuk prediksi konsumsi BBM perusahaan.

UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada PT Abacus atas dukungan fasilitas, sarana, dan prasarana yang telah disediakan, sehingga seluruh rangkaian kegiatan penelitian mulai dari tahap Wawancara sampai dengan perancangan, dapat dilaksanakan dengan baik dan terarah. Penulis juga menyampaikan apresiasi mendalam kepada dosen pembimbing yang telah memberikan arahan, bimbingan ilmiah, masukan konstruktif, serta dukungan moral selama proses penyusunan dan penyelesaian penelitian ini, sehingga penelitian ini dapat terselesaikan sesuai dengan tujuan yang telah ditetapkan.

REFERENSI

- [1] P. A. Septio, S. Yulianto, and J. Prasetyo, "Pembuatan Aplikasi Validasi Document Tagihan Pembelian Barang Secara Digital Menggunakan *OCR* dengan tool

- Tesseract* pada System Portal Perusahaan,” 2023.
- [2] Latifa Khoirani, Rino Ariansyah, and Supiyandi Supiyandi, “Implementasi Algoritma *OCR* (Optical Character Recognition) untuk Ekstraksi Informasi dari Citra Struk Transaksi Keuangan,” *Mars : Jurnal Teknik Mesin, Industri, Elektro Dan Ilmu Komputer*, vol. 2, no. 6, pp. 151–160, Dec. 2024, doi: 10.61132/mars.v2i6.537.
- [3] I. R. Ardiansyah and J. Iskandar, “Sistem Monitoring Volume Bahan Bakar Minyak Untuk Kendaraan Bermotor Berbasis Internet of Things (IoT) Menggunakan Sensor Ultrasonik HY-SRF05 (Studi Kasus Pada Perusahaan Travel Narashansa Transportation),” 2024. [Online]. Available: <https://www.jurnal.stkipppgritulungagung.ac.id/index.php/joincos>
- [4] “Sistem Pengolahan Laporan Data Penjualan BBM Kendaraan pada (SPBU) 24-345-135 dengan Metode System Development Life Cycle.”
- [5] “IMPLEMENTASI ALGORITMA YOLO DAN *TESSERACT OCR* PADA”.
- [6] A. Sonita, “Aplikasi Pendeteksi Obat dan Makanan Menggunakan *OCR* (Optical Character Recognition),” 2018.
- [7] S. P. Apriliando, E. Yulsilviana, J. Daud, W. C. Dharma, and J. M. Yamin, “SISTEM MONITORING DISTRIBUSI SOLAR INDUSTRI PADA PT. SUKSES SANJAYA ENERGY BERBASIS ANDROID.”
- [8] M. Rifky Reyvansyah, E. Setiawan, R. Indarti, I. Munadhif Program Studi Teknik Otomasi, and J. Teknik Kelistrikan Kapal Politeknik Perkapalan Negeri Surabaya, “Penerapan Metode *Optical Character Recognition (OCR)* Untuk Mengambil Data Arsip,” Edy Setiawan, 2022. [Online]. Available: <https://journal.trunojoyo.ac.id/triac>
- [9] K. Oswald Mokorimban, S. Arni, and S. Profesional Makassar, “Monitoring Pengolahan Data Pengguna Bbm Pada Kendaraan Operasional PT,” 2022.
- [10] “Implementasi *Optical Character Recognition (OCR)* untuk Verifikasi Berkas pada Digital Library Program Studi Informatika Universitas Tanjungpura”.

