

BAB I PENDAHULUAN

1.1 Latar belakang

Perkembangan teknologi saat ini telah banyak memberi pengaruh terhadap perkembangan ilmu pengetahuan. Salah satunya, kecerdasan buatan atau *artificial intelligence* merupakan bagian dari ilmu komputer yang membuat agar mesin (*computer*) dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia. Sistem cerdas (*intelligent system*) adalah sistem yang dibangun dengan menggunakan teknik-teknik *artificial intelligence*. Salah satu yang dipelajari pada kecerdasan buatan adalah dalam hal pengenalan pola (*pattern recognition*). Pengenalan pola merupakan suatu ilmu untuk mengklasifikasi atau menggambarkan sesuatu berdasarkan pengukuran kuantitatif fitur atau sifat utama dari suatu objek.

Bagi suatu instansi pemerintah pengenalan pola merupakan suatu sistem pendukung yang memegang peranan penting dalam menunjang kelancaran kegiatan administrasi pada kantor Sekretariat DPRD Provinsi Kepulauan Bangka Belitung terutama mengenai pengarsipan lembaran berita acara serah terima inventaris. Masalah yang dihadapi di Sekretariat DPRD Provinsi Kepulauan Bangka Belitung ini banyaknya arsip yang mesti disimpan kedalam bentuk teks didalam komputer dan pencarian kembali berita acara fisik membutuhkan waktu yang lama. sehingga diperlukan suatu sistem yang dapat membaca dan menyetik otomatis isi dari berita acara tersebut kedalam bentuk digital.

Pengenalan objek (*object recognition*) adalah salah satu aplikasi yang saat ini banyak berkembang dan merupakan aplikasi yang penting dalam dunia pengenalan citra. Aplikasi yang paling banyak dijumpai adalah *Optical Character Recognition* (OCR). Aplikasi OCR sering digunakan untuk mengidentifikasi citra huruf untuk kemudian diubah kedalam bentuk file tulisan (teks) Baik pengenalan plat mobil dari kamera, atau dokumen tulisan tangan yang harus dikonversi menjadi salinan digital, teknik ini sangat berguna. Meskipun tidak selalu sempurna, itu sangat nyaman dan membuatnya jauh

lebih mudah dan lebih cepat bagi beberapa orang untuk melakukan pekerjaan mereka.

Pada penelitian ini, penulis akan mempelajari kedalaman pengenalan karakter optik dan area penerapannya. Penulis juga akan membuat skrip sederhana dengan Python yang akan membantu penulis mendeteksi karakter dari gambar. Pengenalan karakter optik melibatkan pendeteksian konten teks pada gambar dan terjemahan gambar ke teks yang disandikan yang mudah dipahami oleh komputer. Gambar yang mengandung teks dipindai dan dianalisis untuk mengidentifikasi karakter di dalamnya. Setelah identifikasi, karakter tersebut dikonversi menjadi teks yang disandikan dengan mesin. Bagi penulis, teks pada gambar mudah dilihat dan dapat mendeteksi karakter dan membaca teks, tetapi bagi komputer, itu semua adalah serangkaian titik. Gambar dipindai pertama kali dan elemen teks dan grafik diubah menjadi bitmap, yang pada dasarnya adalah matriks titik-titik hitam dan putih. Gambar kemudian diproses terlebih dahulu di mana kecerahan dan kontras disesuaikan untuk meningkatkan akurasi proses. Gambar sekarang dibagi menjadi zona yang mengidentifikasi bidang yang diminati seperti di mana gambar atau teks berada dan ini membantu memulai proses ekstraksi. Area yang mengandung teks sekarang dapat dipecah menjadi garis-garis dan kata-kata dan karakter dan sekarang perangkat lunak dapat mencocokkan karakter melalui perbandingan dan berbagai algoritma deteksi. Hasil akhir adalah teks pada gambar yang kami berikan.

Prosesnya mungkin tidak 100% akurat dan mungkin memerlukan campur tangan manusia untuk memperbaiki beberapa elemen yang tidak dipindai dengan benar. Koreksi kesalahan juga dapat dicapai dengan menggunakan kamus atau bahkan *Natural Language Processing* (NLP). Output sekarang dapat dikonversi ke media lain seperti dokumen kata, PDF, atau bahkan konten audio melalui teknologi text-to-speech. Sebelumnya, digitalisasi dokumen dicapai dengan mengetik teks secara manual di komputer. Melalui OCR, proses ini menjadi lebih mudah karena dokumen dapat dipindai, diproses dan teks diekstraksi dan disimpan dalam bentuk yang dapat diedit seperti dokumen kata. Jika anda memiliki pemindai dokumen di ponsel anda, seperti *Adobe Scan*, Anda mungkin

telah menemukan teknologi OCR sedang digunakan. Bandara juga dapat menggunakan OCR untuk mengotomatiskan proses pengakuan paspor dan ekstraksi informasi dari mereka. Kegunaan lain dari OCR termasuk otomatisasi proses entri data, deteksi, dan pengenalan pelat nomor mobil.

Penulis akan menggunakan pustaka *Python-Tesseract*, atau hanya *PyTesseract*, yang merupakan pembungkus untuk Mesin *Tesseract-OCR Google*. Penulis memilih ini karena sepenuhnya *open-source* dan sedang dikembangkan dan dikelola oleh raksasa yaitu *Google*. Penulis juga akan menggunakan kerangka kerja *Flask* untuk membuat server OCR sederhana kami tempat kami dapat mengambil gambar melalui webcam atau mengunggah foto untuk tujuan pengenalan karakter. Penulis juga akan menggunakan *Pipenv* karena ia juga menangani pengaturan lingkungan virtual dan manajemen persyaratan. Selain itu, penulis juga akan menggunakan pustaka bantal yang merupakan garpu dari *Python Imaging Library (PIL)* untuk menangani pembukaan dan manipulasi gambar dalam banyak format dalam *python*.

Beberapa penelitian terdahulu yang sudah dilakukan terkait dengan penerapan *Optical Character Recognition (OCR)* diantaranya adalah Penelitian yang dilakukan Meland Tetumanto Neno, dkk^[1] pada tahun 2016 yang berjudul “Model Pengenalan Aksara Jepang Menggunakan *Augmented Reality* dan *Optical Character Recognition*” menghasilkan bahwa Pembuatan aplikasi pengenalan karakter aksara Jepang memanfaatkan pustaka *Tesseract* tidak 100% akurat dalam melakukan pengenalan karakter fisik. Fitur *augmented reality* pada aplikasi ini dapat membantu pengguna untuk memahami karakter yang tercetak pada media cetak dan lebih menarik minat pengguna mempelajari aksara Jepang. Penelitian yang dilakukan Samuel ChristianA., dkk^[2] pada tahun 2015 yang berjudul “Sistem Pengaturan Lahan Parkir dengan Metode *Optical Character Recognition (OCR)* Studi Kasus Lahan Parkir Mall Sutos Surabaya” menghasilkan kesimpulan Tingkat akurasi dari sistem cukup rendah dikarenakan sistem tidak menerapkan metode dalam pencarian plat nomor dan pemotongan karakter. Plat nomor dengan kondisi yang kurang baik seperti, pencahayaan pada plat yang tidak merata, cat pada plat terkelupas serta adanya baut plat sehingga mengenai suatu karakter

angka dan huruf plat nomor tersebut yang mengakibatkan tidak terbaca oleh sistem yang dibuat. Tingkat kesadaran dan tanggung jawab dari karakter orang sulit untuk di perhitungkan / diprediksi, sehingga dapat terjadi *human error* pada aplikasi ini dalam pengaturan perpakiran. Penelitian yang dilakukan Muhtadii, Hilmy A. Tawakal^[3] pada tahun 2016 yang berjudul “Pengembangan Aplikasi Android Untuk Pengenalan Citra Nomor Sertifikat Halal MUI dengan *Library Tesseract Optical Character Recognition (OCR)*” menghasilkan kesimpulan bahwa Tesseract OCR kurang cocok bila diimplementasikan pada aplikasi pengenalan citra nomor sertifikat halal, karena nomor sertifikat halal pada kemasan produk yang beredar dipasaran saat ini sebagian besar menggunakan ukuran *font* yang sangat kecil. Penelitian yang dilakukan Sangsaka Wira Utama, dkk^[4] pada tahun 2017 yang berjudul “Aplikasi Pendeteksi Plat Nomor Negara Indonesia Menggunakan *OpenCV* Dan *Tesseract OCR* pada *Android Studio*” menghasilkan kesimpulan bahwa masih diperlukan pemrosesan citra untuk meminimalisir *noise* yang terjadi pada citra untuk meningkatkan akurasi konversi citra menjadi karakter. Penelitian yang dilakukan Darmawan Aditama, dkk^[5] pada tahun 2018 yang berjudul “Aplikasi Pembelajaran Alfanumerik Untuk Anak Usia Pra-Sekolah Berbasis Android Menggunakan Metode *Tesseract-Ocr*” menghasilkan kesimpulan bahwa Tesseract dengan bantuan *library Tess-two* dapat mengenali tulisan tangan dengan bentuk alphabet kapital dan numeric yang ditulis oleh pengguna aplikasi. Namun aplikasi tidak berjalan dengan baik pada smartphone android yang memiliki resolusi lebih besar dari HVGA (320X480 Pixels), sehingga masih perlu dikembangkan lagi untuk dapat dijalankan pada smartphone android dengan resolusi yang lebih besar.. Penelitian yang dilakukan Fauzan Adiima, dkk^[6] pada tahun 2018 yang berjudul “Pembangunan Perangkat Lunak Bantu untuk Menyelesaikan Persoalan Matematika dengan Menggunakan *Optical Character Recognition* Berbasis Android” menghasilkan kesimpulan dari hasil pengujian dari bab empat, pada skenario pengujian karakter dan simbol yaitu 71.42% terbaca, 14.28% sebagian terbaca dan 0% yang tidak terbaca. Lalu pada skenario pengujian perhitungan yaitu mampu menjawab soal pengujian sebanyak 20 dengan perhitungan yang benar. Sedangkan untuk skenario pengujian tulisan

tangan adalah 62.85% terbaca, 20% sebagian terbaca dan 2.85% yang tidak terbaca dengan waktu kurang lebih 1 detik dari setiap pengujian.

Berdasarkan penjabaran diatas, penulis tertarik mengambil judul penelitian **“Model Pengenalan Teks Pada Lembaran Berita Acara Serah Terima Inventaris Pada Kantor DPRD Provinsi Kepulauan Bangka Belitung”**.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah disampaikan, maka rumusan masalahnya adalah “Bagaimana membuat model pengenalan teks pada lembaran berita acara serah terima inventaris pada Kantor Sekretariat DPRD Provinsi Kepulauan Bangka Belitung?”

1.3 Batasan Masalah

Batasan masalah dari penelitian ini sebagai berikut:

1. Pengolahan citra yang digunakan *Tesseract* sebagai mesin pengenalannya.
2. Penelitian ini hanya membahas mengenai pengenalan teks untuk dimanfaatkan sebagai pengarsipan lembar berita acara serah terima inventaris pada kantor Sekretariat DPRD provinsi bangka belitung.
3. Output dari hasil pengenalan teks dalam bentuk file .html.
4. Inputan pada aplikasi pengenalan teks ini berupa file jpg, gif, tif dan png.

1.4 Tujuan dan Manfaat

Penelitian ini bertujuan untuk membuat model pengenalan teks pada lembaran berita acara serah terima inventaris pada kantor Sekretariat DPRD provinsi kepulauan bangka belitung

Adapun penelitian ini memberikan manfaat bagi Sekretariat DPRD untuk mengatasi permasalahan pembuatan dokumen ulang untuk lembaran berita acara serah terima inventaris, karena penyimpanan arsip yang tidak rapi atas filenya sehingga saat diperlukan tidak perlu dibuat ulang.pembuatan dokumen

pengarsipan disimpan dalam bentuk digital agar mudah melakukan penelusuran dikemudian hari.

1.5 Sistematika Penulisan Laporan

Gambaran yang lengkap dan jelas mengenai masalah yang akan dibahas penulis dibagi menjadi beberapa bagian sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi penjelasan tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisannya.

BAB II LANDASAN TEORI

Bab ini merupakan tinjauan pustaka yang mengurai teori-teori yang mendukung judul dan mendasari pembahasan secara rinci. Pada bab ini juga dituliskan tentang alat bantu pengembangan sistem yang digunakan untuk pengembangan pada penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang model pengembangan perangkat lunak, metode pengembangan perangkat lunak, dan *tools* yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang struktur organisasi jabatan tugas dan wewenang, analisis masalah sistem yang berjalan, analisis kebutuhan sistem usulan, analisis sistem, dan perancangan sistem.

BAB V PENUTUP

Pada bab ini berisi penjelasan tentang kesimpulan dan saran yang disimpulkan peneliti setelah merancang, menguji, dan menjalankan aplikasi, serta saran yang ditujukan untuk pengembang