BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan pengujian sistem ekstraksi data otomatis dari citra Kartu Tanda Mahasiswa (KTM) menggunakan Optical Character Recognition (OCR) dan Regular Expression berbasis GUI Tkinter yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut.

1. Sistem yang dibangun berhasil mengotomatiskan proses ekstraksi lima field data (NPM, Nama, Program Studi, Fakultas, dan TTL) dari citra KTM melalui *pipeline* yang terdiri atas prapemrosesan citra (*resize*, *deskew* dengan penjagaan, *crop* zona teks, dan binarisasi Otsu), pengenalan teks menggunakan Tesseract OCR dengan strategi multi-sumber dan multi-PSM disertai optimasi *early-exit*, serta ekstraksi field terstruktur menggunakan pola Regex bertingkat dengan mekanisme *fallback*.
2. Penyederhanaan *pipeline* prapemrosesan citra dengan menghilangkan CLAHE dan *sharpening kernel* yang pada rancangan awal justru merusak keterbacaan teks pada citra berlatar belakang bertekstur, serta menggantinya dengan binarisasi Otsu global — terbukti secara signifikan meningkatkan keterbacaan hasil OCR dibandingkan pendekatan prapemrosesan yang lebih agresif.
3. Hasil pengujian terhadap 100 citra KTM menunjukkan sistem mencapai akurasi Overall sebesar 93,8%, dengan rincian akurasi NPM 100%, Fakultas

98%, Program Studi 95%, TTL 94%, dan Nama 82%, yang membuktikan bahwa pendekatan OCR dan Regex mampu mengekstraksi data KTM secara akurat pada mayoritas kasus tanpa memerlukan model *machine learning* yang kompleks.

5.1 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan, berikut beberapa saran untuk pengembangan lebih lanjut.

1. Field Nama masih memiliki akurasi terendah di antara kelima field yang diekstraksi (82%), sehingga penelitian selanjutnya dapat mengeksplorasi mekanisme pembersihan token yang lebih adaptif atau pemanfaatan basis data nama untuk memvalidasi hasil ekstraksi, tanpa mengorbankan risiko memotong nama pendek yang sah.
2. Tahap prapemrosesan citra pada penelitian ini menggunakan binarisasi Otsu global yang terbukti efektif meningkatkan keterbacaan OCR dibandingkan pendekatan CLAHE dan sharpening yang lebih agresif, namun efektivitas ini baru diuji pada kondisi pencahayaan dan latar belakang templat KTM UMB. Penelitian selanjutnya disarankan menguji efektivitas kombinasi teknik prapemrosesan lain, misalnya adaptive thresholding atau image enhancement berbasis deep learning, khususnya untuk menangani citra dengan pencahayaan tidak merata, bayangan, atau kualitas kamera yang lebih rendah.
3. Sistem saat ini bergantung sepenuhnya pada mesin OCR Tesseract berbasis aturan dan pola. Penelitian lanjutan dapat mengeksplorasi penggunaan model OCR berbasis *deep learning* atau model *Vision-Language* untuk menangani

citra dengan kualitas rendah atau kemiringan ekstrem yang sulit ditangani oleh pendekatan berbasis Regex.

4. Waktu pemrosesan batch, meskipun telah dipercepat melalui strategi *early-exit*, masih dapat dioptimalkan lebih lanjut misalnya melalui pemrosesan paralel (*multiprocessing*) pada pengujian data dalam jumlah besar