

ABSTRAK

Optical Character Recognition (OCR) merupakan teknologi yang memungkinkan komputer untuk mengenali dan mengonversi teks pada citra digital menjadi data teks yang dapat diolah lebih lanjut. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Long Short-Term Memory* (LSTM) pada sistem OCR untuk mengekstraksi teks dari citra soal buku sekolah Bahasa Indonesia kelas X dan mengevaluasi tingkat akurasi. Sistem yang dibangun terdiri dari dua tahap utama, yaitu pelatihan model LSTM dan pengujian model terhadap citra dokumen soal. Proses pelatihan dilakukan menggunakan 2.342 data latih dengan 73 jenis karakter, sedangkan pengujian dilakukan pada 30 data uji hasil *scan* buku soal dengan *font Times New Roman*. Hasil pengujian menunjukkan tingkat akurasi rata-rata sebesar 88,50% dengan akurasi tertinggi sebesar 92,89%. Selain itu, dilakukan validasi eksternal dengan melibatkan lima guru SMA sebagai responden untuk mengevaluasi hasil OCR berdasarkan aspek kesesuaian hasil, keterbacaan teks, dan kelayakan penggunaan. Berdasarkan hasil kuesioner, sistem dinilai layak digunakan dengan skor rata-rata kesesuaian 4 (Setuju), keterbacaan 3,8 (Cukup Setuju hingga Setuju), dan kelayakan penggunaan 4,2 (Setuju). Sistem ini bermanfaat untuk mempercepat proses digitalisasi dokumen soal dan mendukung efisiensi guru dalam mengelola bank soal digital. Namun, akurasi pengenalan karakter masih dapat ditingkatkan melalui optimasi pada tahap segmentasi untuk mengatasi kasus karakter berdempet dan noise pada citra input. Penelitian ini memberikan kontribusi awal dalam pengembangan teknologi OCR berbasis *deep learning* di bidang pendidikan.

Kata kunci — OCR, LSTM, pengenalan karakter, pengolahan citra.