

ABSTRAK

Aksara daerah merupakan sistem ortografi hasil masyarakat daerah yang meliputi aksara dan sistem pengaksaraan untuk menuliskan bahasa daerah, salah satunya adalah Aksara Sunda. Berbagai penelitian sebagai usaha digitasi dalam pelestarian Aksara Sunda dengan berbagai metode dalam *machine learning* dengan fokus *Optical Recognition Character* (OCR) terutama *Convolutional Neural Network* (CNN) sudah dilakukan. Namun, data yang ada pada saat ini terbatas pada karakter yang menjadi bagian swara, ngalagena, rarangkén dan angka. Masih sedikit sistem OCR yang menggunakan dataset kata Aksara Sunda yang berupa gabungan antara ngalagena dan swara dengan rarangkén. Dengan penambahan dataset kata tersebut dapat menghasilkan dua model, yaitu model *ensemble* dan model *Connectionist Temporal Classification* (CTC). Model *ensemble* dengan *ensemble weight* sebesar 0,5 berhasil mencapai akurasi sebesar 0,912, presisi sebesar 0,889, sensitifitas sebesar 0,701, dan *F1-score* sebesar 0,755. Model CTC berhasil mencapai nilai *Character Error Rate* (CER) sebesar 0,069 dan nilai *Word Error Rate* (WER) sebesar 0,087.

Kata Kunci: *Optical Character Recognition*, Aksara Sunda, *Convolutional Neural Network*, *Recurrent Neural Network*, *Connectionist Temporal Classification*, *EfficientNet*.