

BAB III

METODOLOGI

3.1 Jenis Penelitian

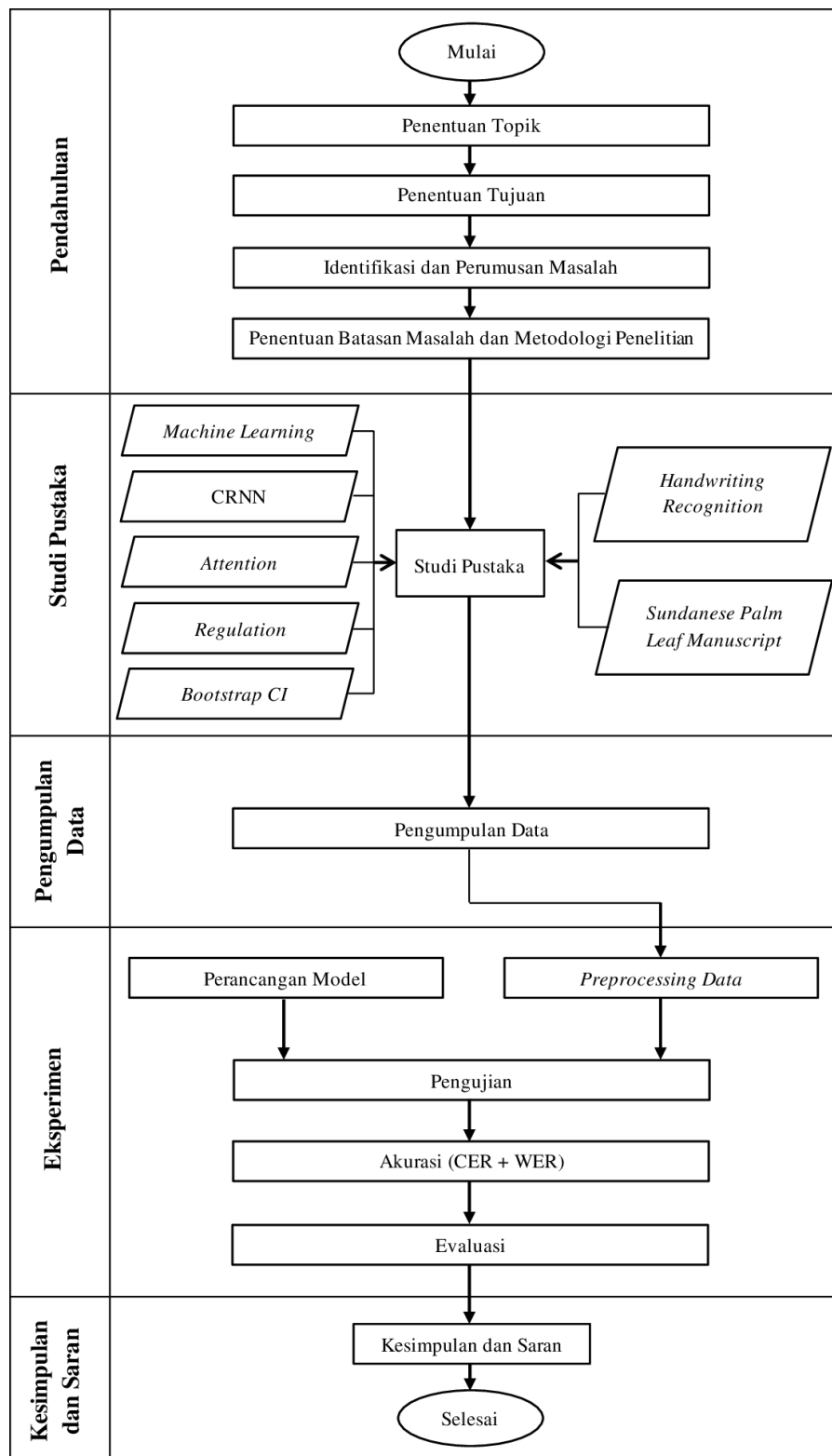
Jenis penelitian yang digunakan merupakan jenis eksperimental. Jenis eksperimental digunakan untuk mengetahui pengaruh dalam pemberian atau perubahan suatu variabel terhadap objek penelitian, dalam hal ini variabel berupa parameter diuji terhadap model eksperimen yang dibuat berupa CRNN dengan Attention untuk mendapatkan tingkat *word error rate* pada pengenalan manuskrip lontar Sunda kuno.

3.2 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini berupa data sekunder dari ICHFR 2018 *Challenge D* yang bersumber dari ICDAR 2017 (Suryani, Paulus, Hadi, Darsa, & Burie, 2017), data terdiri dari gambar dan label yang telah tersegmentasi perkata. Selain data sekunder penelitian ini melibatkan data buatan berupa data *mapping* yang terdiri dari simbol (termasuk *null*), angka, huruf kecil, dan huruf kapital dengan total sebanyak 80 karakter yang digunakan untuk representasi label dalam bentuk angka maupun sebaliknya.

3.3 Tahapan Penelitian

Tahapan yang dilakukan pada penelitian ini dapat dijabarkan sebagai berikut:



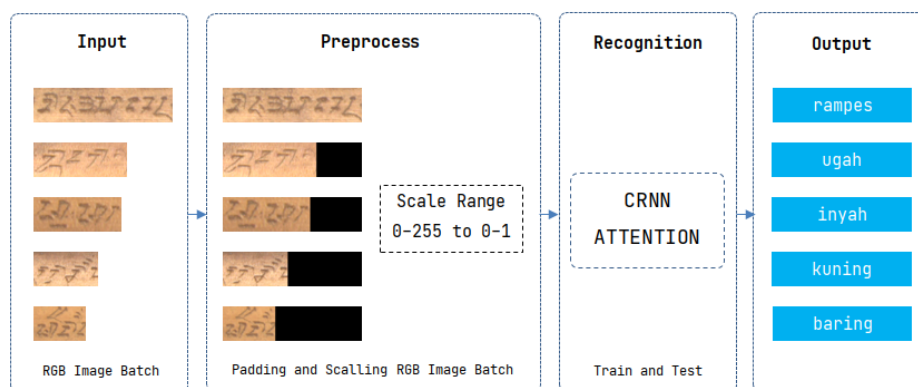
Gambar 3.1 Tahapan Penelitian

3.4 Tahapan Eksperimen

Eksperimen ini dibagi menjadi beberapa tahapan sebagai berikut:

1. Perancangan Model

Model yang akan dibuat, diproses dengan rancangan sistem sebagai berikut:



Gambar 3.2 Arsitektur Sistem

Secara runtut dan umum sistem terdiri dari *input* yang merupakan data masukkan untuk model dengan *preprocess* input terlebih dahulu, kemudian input diproses oleh model CRNN + *Attention* sehingga menghasilkan output berupa teks yang diprediksi.

Model CRNN + *Attention* pada penelitian ini menggunakan fitur ekstraksi (Puigcerver, 2017) sebagai dasarnya dengan penambahan operasi *layer* pada akhir CNN berupa operasi *reshape* serta *adaptive average pooling layer*, untuk varian RNN yang digunakan berupa 1 *layer Bidirectional LSTM* dengan 100 *hidden size* dan menggunakan *Linear layer* untuk *mapping* terhadap *class* serta LogSoftmax untuk perhitungan probabilitas prediksi.

2. *Preprocessing data*

Preprocessing data merupakan proses normalisasi data yang bertujuan agar data dapat diproses dengan baik oleh model baik dalam ruang lingkup *training* (pelatihan) atau *testing* (pengujian) model. Data yang akan digunakan untuk pelatihan dan pengujian dilakukan *padding* berdasarkan panjang gambar terpanjang pada *batch* serta penskalaan nilai (*array image*) RGB dari rentang 0-255 menjadi 0-1.

3. Pengujian (*training* dan *testing*)

Model yang telah dirancang dan telah diimplementasi dalam bentuk yang dapat dieksekusi dilatih sesuai dengan *parameter* yang telah diatur, metrik yang diukur pada pelatihan berupa *loss*, *character error rate* (CER), dan *word error rate* (WER) dalam setiap *epoch*. Pengujian model yang dilakukan pada eksperimen ini dilakukan untuk setiap *epoch*, dimana satu *epoch* merupakan pelatihan dari keseluruhan data *train*. Metrik yang dievaluasi berupa CER, WER pada data *test*.

4. Evaluasi

Kumpulan CER, WER, serta *loss* yang didapat setelah pengujian secara menyeluruh dievaluasi sebagai bahan untuk penarikan kesimpulan.