**Bab III**

**Perancangan Pengujian**

## 3.1 Perancangan Awal Sistem dan Eksperimen Pengenalan Huruf Arab

Dalam tesis ini akan dibuat sebuah perangkat lunak pengenalan huruf Arab yang bisa mengenali huruf Arab terisolasi dan dalam rangkaian kata. Masukan dari sistem ialah berupa citra berisikan kalimat tulisan Arab kemudian melalui proses binerisasi. Untuk pengenalan huruf Arab terisolasi setelah melalui proses binerisasi akan dilanjutkan ke tahap penipisan, kemudian fitur dari tiap huruf akan di ekstrak melalui chain code, jumlah titik dan posisi titik. Setelah fitur di ekstrak akan dilakukan klasifikasi menggunakan neural network dan hidden markov model. Sedangkan pada pengenalan huruf Arab dalam kalimat setalah proses binerisasi akan melalui tahap segemntasi, kemudian dilanutkan pada tahap penipisan, lalu akan di ekstrak fitur setiap huruf dengan chain code, umlah titik dan posisi titik. Setelah fitur di ekstrak akan dilakukan klasifikasi menggunakan neural network dan hidden markov model. Gambaran umum tahapan proses pengenalan huruf Arab dapat dilihat pada Gambar III.1



Gambar III.1 Alur proses pengenalan huruf Arab

Gambar III. 1 Diagram blok utama dalam sistem pengenalan huruf Arab

Dalam penelitian ini dikenal data latih dan data uji. Berikut penjelasan dari kedua jenis data tersebut.

1. Data latih

Data latih merupakan kumpulan data huruf arab sesuai sesuai posisinya pada kalimat yakni terisolasi, di awal, di tengah, dan di akhir yang akan dilatih untuk kemudian disimpan data latih huruf Arab.

1. Data uji

Data uji terbagi dua, yaitu data uji huruf Arab terisolasi dan kumpulan data kalimat Arab yang akan dimasukan dalam tahap pengujian dan dibandingkan data latih untuk ditentukan hasil pengenalan huruf Arab tersebut.

Jumlah huruf arab yang akan dilatih sebanyak 31 huruf arab, 1 huruf arab memiliki 4 posisi yang berbeda-beda, yaitu posisi terpisah, diawal, ditengah dan diakhir. Berikut teknik yang digunakan dalam pengambilan data latih.

1. Ketik huruf 31 huruf Arab menggunakan Microsoft Word , huruf arab ditulis sesuai posisinya pada kalimat, yakni posisi terisolasi, di awal, di tengah dan di akhir.
2. Potong setiap huruf menggunakan Snipping Tool
3. Hasil pemotongan huruf disimpat sebagai data latih huruf arab

Saat pengambilan data uji, dilakukan dalam dua tingkat kesulitan, yaitu:

1. Pengujian huruf Arab terisolasi

Saat pengujian huruf ini dilakukan sebanyak 31 huruf arab terisolasi yang terdapat pada tabel II.1. Pangujian huruf Arab terisolasi dilakukan dalam waktu 13 detik.

1. Pengujian kalimat

Saat pengujian kalimat dilakukan dengan memasukkan 10 kalimat yang berbeda kedalam sistem. Pengujian kalimat dilakukan dalam waktu 11 detik. Berikut 10 kalimat Arab yang akan diujikan kedalam sistem :

1. الاحترام المتبادل بين الاديان
2. لقمر تبدو جميلة جدا
3. ف كهة دوريان ا لا ذو اق جيدة
4. تتخلي ابدا في الحياة
5. انتظر اي محا كمة
6. لحفاظ علي صحتك حياة طيبة
7. المعلمين يعلمو ن
8. جدةالارز المطبو خ
9. يتكلم ببطء مفهومة حتي
10. الحفا ظ علي صحتك حيا ة طيبة

**III.2 Perancangan Sistem dan Eksperimen**

Pada penelitian ini terdapat 5 tahap yang dilakukan untuk membuat sistem pengenalan huruf Arab, berikut adalah penjelasan masing – masing tahap.

**III.2.1 Eksperimen Binerisasi**

Binariisasi gambar adalah proses pengubahan gambar menjadi biner yang memiliki nilai 0 dan 1. Citra grayscale akan berubah menjadi hitam putih. Proses binerisasi diperlukan untuk melakukan langkah selanjutnya pada pengenalan huruf dan kalimat Arab. Cara melakukannya adalah dengan melakukan threshold pada setiap channel warna. Ambang yang digunakan adalah 150. Jika saluran warna kurang dari 150 maka akan dikonversi menjadi hitam, dan jika warnanya lebih dari 150 akan berubah menjadi putih.



Sebelum Binerisasi Sesudah Binerisasi

Gambar III.2 Perbandingan citra sebelum binerisasi dan sesudah binerisasi

**III.2.2 Perancangan Sistem Penipisan**

Setelah diperoleh citra biner, dilakukan proses penipisan untuk mendapatkan tulang dari huruf. Proses penipisan dilakukan dengan algoritma Stentiford. Langkah – langkah dari algoritma Stentiford telah dijelaskan pada bab II.4.2.

**III.2.3 Eksperimen Segmentasi**

Pada pengenalan huruf Arab dalam kalimat melalui tahap segmentasi terlebih dahulu sebelum tahap penipisan.

Segmentasi dilakukan dengan menggunakan algoritma Zidouri [7]. Langkah pertama segmentasi adalah menentukan beberapa parameter yang digunakan sebagai segmen segmentasi. Setelah parameter dilakukan tahap segmentasi yang dilakukan. Kemudian akan dipilih guide band sebagai referensi segmentasi karakter. Untuk memilih panduan band yang benar beberapa fitur diekstraksi dari masing-masing guide band.

Ada empat aturan yang digunakan algoritma saat ini calon band giude:

1. Pilih guide band jika lebar relatifnya paling besar dan F4 = 1
2. Pilih guide band jika F2 > Ls dan F4 = 1
3. Pilih guide band jika F2 < Ls dan F3 > Lsa dan pita bukan pita terakhir
4. Pilih guide band jika F1 lebih dari Lm dan F4 = 1

dengan F1 = lebar guide band, F2 = jarak guide band ke guideband sebelah kanannya, F3 = jarak guide band ke guide band kedua sebelah kanannya, dan F4 = posisi ditemukannya guide band, bernilai satu di atas lini basis dan nol jika di bawah (Zidouri, 2010).

**III. 2. 4 Eksperimen Ekstraksi Fitur**

Ekstraksi fitur dilakukan dalam 3 tahap yakni sebagai berikut :

1. Chain code

Setelah melalui tahap penipisan fitur chain code di ekstrak untuk setiap huruf Arab kemudian dialkukan normalisasi pada setiap chain code sehingga panjang chain code hanya 10 untuk setiap huruf. Langkah – langkah untuk mengambil chain code pada suatu objek telah dijelaskan pada bab II.4.1, dan langkah – langkah untuk melakukan normalisais chain code telah dijelaskan pada tahap II.4.1.1

1. Jumlah titik

Setelah fitur chaincode fitur jumlah titik merupakan fitur yang penting karena beberapa huruf arab memiliki kemiripan bentuk namun dapat di bedakan berdasarkan jumlah titik, jumlah titik terbagi menjadi 4 , yakni 0, 1, 2 dan 3.

1. Posisi titik

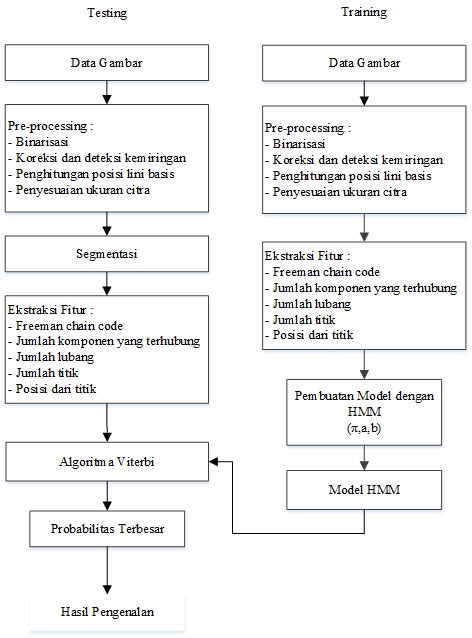
Setelah fitur chaincode dan jumlah titik, fitur posisi titik juga merupakan fitur yang penting karena beberapa huruf arab memiliki kemiripan bentuk namun dapat di bedakan berdasarkan posisi titik, posisi titik terbagi menjadi 3, yaitu di atas, ditengah dan di bawah.

Sebelum diproses lebih lanjut citra akan melewati beberapa tahap praolah yakni sebagai berikut.

**3.5 Perancangan Sistem dan Eksperimen Pengklasifikasi**

**3.5.1 Tahap Pelatihan dan Pengujian dengan Metode Hidden Markov Model**

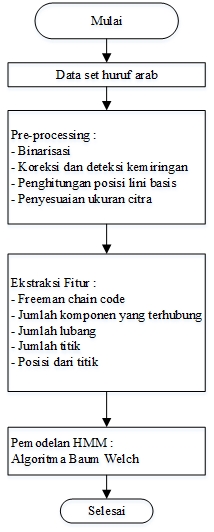
Dalam penelitian ini menggunakan Hidden Markov Model (HMM) untuk tahap klasifikasi. Secara umum, diagram blok dari proses pengenalan huruf arab menggunakan metode Hidden Markov Model dalam Gambar 3.2.



Gambar 3.2 Proses pengenalan huruf arab menggunakan Hidden Markov Model

1. Tahap pelatihan dengan metode Hidden Markov Model

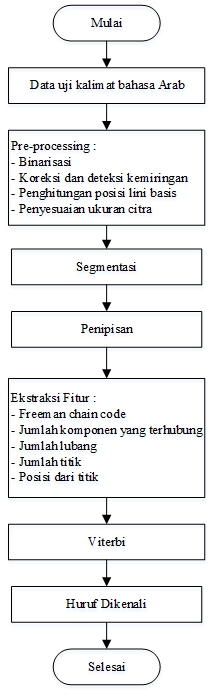
Tahap pelatihan dilakukan untuk menentukan parameter estimasi, sehingga  
terbentuk model markov tersembunyi berupa . Berikut diagram alir pelatihan dengan metode model markov tersembunyi.



Gambar 5.11 Diagram alir pelatihan dengan metode *Hidden Markov Model* (HMM)

1. Tahap pengujian dengan metode model markov tersembunyi

Pengujian dilakukan untuk mengetahui seberapa besar nilai peluang data citra uji  
terhadap kecocokannya dengan data latih*.* Proses pengujian atau evaluasi ini menggunakan algoritma *viterbi* karena yang ingin dicari hanya nilai peluang dari data uji di setiap basis data hasil pelatihan. Nilai peluang yang paling tinggi, menentukan kecocokan antara data uji terhadap basis data latih. Berikut algoritma *viterbi* yang digunakan dalam tahap pengujian dengan metode HMM.



Gambar 5.12 Diagram alir pengujian dengan Hidden Markov Model

Tahapan pertama adalah memasukkan gambar rangkaian kata dalam huruf Arab. Terhadap gambar tersebut dilakukan pemrosesan awal guna membersihkan gambar huruf dari kerusakan atau karena posisi yang tidak sesuai serta mengeliminasi hal-hal yang tidak diperlukan. Warna gambar kemudian diubah ke dalam warna abu- *(gray scale)*, dan selanjutnya diubah lagi ke dalam warna hitam-putih, selanjutnya dilakukan proses *thinning* dengan algoritma *Zhang Suen* yang berguna untuk mendapatkan jalur huru, kemudian dilakukan tahapan segmentasi dari kalimat menjadi huruf. Huruf yang telah berhasil di segmentasi dimasukkan kedalam metode *Freeman Chain code* dalam tahap ekstraksi fitur pengambilan ciri masing-masing huruf. Hasil dari ekstraksi fitur ini akan digunakan sebagai masukan untuk tahap klasifikasi dengan *Hidden Markov Model* (HMM) yang merupakan proses identifikasi masing-masing karakter dan menetapkan kedalam kelasnya sendiri.