

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Banyak generasi muda Bali saat ini kurang mengenal aksara, membaca ataupun menulis Bali (Cahyadhi, Sunarya, & Wirawan, 2015) . Hal ini disebabkan pengaruh globalisasi yang menggerus kearifan lokal. Walaupun orang Bali bisa menggunakan bahasa Bali namun sulit untuk membaca aksara Bali pada lontar (Dharma, 2016). Selain aksara yang tidak terbaca akibat usia lontar, juga tradisi yang menganggap membaca aksara Bali adalah hal tabu, sakral dan angker. Biasanya lontar baru akan ditunjukkan pada hari atau upacara tertentu karena dianggap sebagai pusaka. Lontar hanya diperlihatkan sebagai bagian upacara sehingga jarang dibuka dan dibaca (M. Kesiman & Burie, 2016) .

Penulisan lontar menggunakan pisau kecil (*pengutik* atau *pengropak*) pada lontar (Negari, Kuna, & Pallawa, 2016). Penulisan ini seperti tulisan tangan (*handwritten*). Penulisan aksara Bali pada lontar memiliki tantangan dari hasil penulisan. Misal aksara Bali yang ditulis oleh orang yang sama hasil penulisan bisa berbeda. Perbedaan dalam karakter ini membuat sulit untuk dibaca oleh pembaca awam sehingga tidak mudah ditafsir (Wiguna, 2016). Selain sulit dibaca, penggunaan lontar bersifat jangka pendek karena lontar bisa mengalami penurunan kualitas secara fisik. Penurunan ini bisa berupa warna lontar berubah warna atau rusak dimakan rayap (lihat Gambar 1.1: Contoh Lontar yang Rusak

Dimakan Rayap) . Penurunan kualitas ini menyebabkan aksara di lontar tidak bisa dibaca (Udayana, Sudarma, & Kumara, 2017).



Gambar 1.1: Contoh Lontar yang Rusak Dimakan Rayap

Seperti sebuah buku, lontar memuat hal-hal yang bertemakan pendidikan, informasi, maupun puja dan puji kepada Tuhan. Jika disamakan dengan masa sekarang, lontar sama seperti buku dongeng, majalah, atau buku-buku doa yang dijual bebas di toko buku. Misal Lontar Siwaratri Kalpa yang menceritakan tentang Lubdaka, seorang pemburu yang masuk surga karena bersemadi saat Siwaratri. Lontar Tujur Usada Bali yang diperkirakan ditulis pada tahun 1800an. Lontar ini berisi naskah yang mengandung legenda, mitos, dan sejarah raja-raja dari periode Bali Kuno sampai terbentuknya Kerajaan Gelgel (Prasetya, 2016). Banyaknya lontar yang memiliki informasi untuk digali sehingga perlu diselamatkan dan dilestarikan sebelum rusak akibat umur lontar atau dimakan rayap (Primadesi, 2018). Salah satu upaya untuk pelestarian dengan digitalisasi lontar dengan *Optical Character Recognition* (OCR) atau pengenalan karakter

(Rino, n.d.). OCR merupakan konversi mekanik atau elektronik pada sebuah komputer dari gambar tulisan tangan, ketik, atau cetak ke dalam encode teks (Morse & Patel, 2007).

Penelitian digitalisasi atau OCR aksara Bali pada lontar telah dilakukan (M. W. A. Kesiman, Prum, Burie, & Ogier, 2015). Proses OCR ada beberapa bagian yaitu Praproses, Pengenalan Karakter, dan Pasca-Proses. Praproses yang dilakukan berupa penghalusan tepi, binarisasi (M. W. A. Kesiman et al., 2015), penghapusan garis, analisa layout, segmentasi karakter (M. Kesiman & Burie, 2016). Hasil segmentasi karakter dikenali sebagai *Isolated Handwrittern Character Recognition* (IHCR) (M. W. A. Kesiman, Prum, Burie, & Ogier, 2017)

Pengenalan karakter tulisan tangan (*IHCR*) adalah penelitian yang intensif selama 3 dekade terakhir (Burie et al., 2017) . Berbagai metode *IHCR* telah mencapai kinerja yang memuaskan terutama untuk aksara Latin. Penelitian *IHCR* untuk aksara selain Latin menjadi tantangan utama bagi peneliti. Penelitian *IHCR* naskah daun lontar dari Thailand (Aggarwal, Singh, & Singh, 2015) , dari Sunda (Riansyah, Nurhasanah, & Dewi, 2017) dan dari Bali (M. W. A. Kesiman et al., 2017)

Pada penelitian *IHCR* dari Bali adalah set data dan pengenalan. Penyediaan set data dengan digitalisasi lontar memiliki beberapa tahapan. Tahapan pertama lontar dikumpulkan di berbagai daerah di Bali, diambil gambar lontar setiap lembar menggunakan kamera. Hasil gambar lontar akan dibagi menjadi beberapa kata dengan ALETHEIA (<http://www.primaresearch.org/>)

tools/Aletheia) secara manual. Pembagian kata ini dipecah lebih kecil menjadi karakter dengan metode *Otsu*. Hasil karakter kemudian validitas oleh filolog sehingga menghasilkan *isolated character* dan dijadikan sebagai salah bagian set data *AMADI_LontarSet* (Windu et al., 2016). Pada set data tersebut berjumlah 133 karakter dan memiliki jumlah anggota yang tidak berimbang karena tidak semua karakter tersedia pada seluruh lontar Tabel 2.2: Nama Kelas Karakter dan Jumlah Anggota Kelas di Data Set *AMADI_LontarSet*.

Selanjutnya tahapan pengenalan *isolated character* pada *IHCR*. Pengenalan memiliki dari beberapa langkah seperti fitur ekstraksi dan klasifikasi. Fitur ekstraksi adalah mengekstrak informasi dari lontar dengan tujuan klasifikasi (Aggarwal et al., 2015). Banyak metode ekstraksi fitur dan klasifikasi telah diusulkan untuk pengenalan karakter (Kumar, 2013) (Arica & Yarman-Vural, 2002). Metode tersebut diimplementasikan dan dievaluasi untuk pengenalan karakter India (Ramteke, 2010), Cina (Jin et al., 2009) (Liu, Yin, Wang, & Wang, 2013) , dan Jepang (Kimura & Miyake, 1987). Untuk mendapatkan kinerja sistem yang baik tergantung pada pemilihan ekstraksi fitur dan klasifikasi yang tepat (Kozlay, 1971)

Penelitian terkait pengenalan karakter aksara Bali dengan eksperimen lontar menggunakan set data *AMADI_LontarSet* dengan 10 fitur metode ekstraksi (M. W. A. Kesiman et al., 2017). Penelitian ini mengusulkan kombinasi metode ekstraksi fitur untuk meningkatkan tingkat pengenalan karakter. Hasil penelitian ini menunjukkan tingkat pengenalan karakter dapat meningkat secara signifikan

dengan 4 ekstraksi fitur yang digabungkan. Ekstraksi fitur tersebut *Neighborhood Pixels Weights* dengan *Kirsch (NPWK)* yang dikombinasikan *Histogram of Gradient (HoG)* dan metode *Zoning* kemudian klasifikasi dengan *k-Neighbour Nearest (k-NN)*. Metode gabungan ini menghasilkan tingkat pengenalan hingga 85%. Hasil eksperimen ini masih sedikit lebih baik daripada menggunakan *Convolutional Neural Network (CNN)*. Metode *CNN* memiliki nilai pengenalan 84,30 % akibat tidak berimbang jumlah kelas karakter yang digunakan. Padahal metode *CNN* untuk klasifikasi karakter dengan set data *MNIST* yang berisi citra angka memiliki nilai akurasi yang tinggi.

Metode *Virtual Modified Quadratic Discriminant Function (VMQDF)* mencapai akurasi tertinggi pada kompetisi *Internasional Conference on Frontier In Handwritten Recognition (ICFHR) 2016* (Burie et al., 2017) dengan menggunakan set data *AMADI_LontarSet* untuk tantangan 3 Isolated Character. Metode ini diusulkan Group I yang berisi *Yunxue Shao, Jiantao Zhou and Guanglai Gao*, dari *College of Computer Science, Inner Mongolia University*. Metode ini memiliki nilai akurasi 88,39 % atau lebih baik dibandingkan metode gabungan ekstraksi fitur. Penjabaran Algoritma *VMQDF* seperti berikut 1) membuat virtual sampel set data sebanyak 97 sehingga jumlah kelas karakter menjadi lebih banyak dibandingkan set data. 2) proses *Gradient Feature* dengan metode *Normalization Cooperated Gradient Feature (NCFG)*. 3) klasifikasi dengan *Modified Quadratic Discriminant Function (MQDF)*. *MQDF* adalah metode klasifikasi statistik untuk meningkatkan perhitungan efisiensi dan kinerja

dengan modifikasi QDF melalui *eigenvalue* dan *eigenvector* (Liu, Sako, & Fujisawa, 2004). QDF adalah klasifikasi dengan diskriminasi tiap kelas dari Distribusi *Gaussian* dan menggunakan parameter ke dalam *Bayes* yang dimana tiap kelas memiliki matriks kovarian (Casella, Fienberg, & Olkin, 2006). QDF lebih baik dari segi kompromi dalam penggunaan masukan variabel dibandingkan k-NN. Selain itu k-NN memiliki kelemahan terkait jumlah data pada kelas yang tidak seimbang. Jika jumlah anggota pada kelas lebih banyak maka anggota kelas yang lebih sedikit akan digolongkan sebagai anggota kelas yang lebih banyak (Genesis, 2018).

Berdasarkan permasalahan dan ide di atas terkait metode terbaru dalam pengenalan aksara Bali pada lontar maka muncul gagasan terkait komparasi kinerja klasifikasi. Komparasi yang dilakukan dengan membandingkan klasifikasi *MQDF* dengan *Discriminative Learning Quadratic Discriminant Function* (DLQDF). Metode *MQDF* dan *DLQDF* ini merupakan pengembangan dari *Quadratic Discriminant Function* (QDF). Alasan pemilihan 2 metode ini karena memiliki akurasi yang tertinggi pada pengenalan karakter China (Liu et al., 2013). Pada penelitian ini melakukan komparasi klasifikasi *MQDF* dan *DLQDF* akan dibandingkan k-NN

1.2 Identifikasi Masalah

Maka berdasarkan latar belakang yang telah dijelaskan di atas maka pada penelitian ini dapat dikenali beberapa masalah sebagai berikut:

1. Orang-orang Bali jarang membaca lontar karena bersifat keramat sehingga kesulitan mendapatkan informasi. Oleh karena itu dipandang perlu upaya digitalisasi
2. Para peneliti melakukan digitalisasi dengan menggunakan OCR pada lontar namun untuk pengenalan karakter aksara Bali belum baik
3. Salah satu penelitian sebelum (M. W. A. Kesiman et al., 2017) Untuk pengenalan *isolated character* aksara Bali menggunakan gabungan metode ekstraksi fitur mencapai akurasi 85%. dengan klasifikasi k-NN.
4. Pada penelitian tersebut penggunaan klasifikasi k-NN memiliki kelemahan dengan jarak jumlah anggota kelas. Bila jarak antar anggota kelas yang kecil akan lebih digolongkan sebagai anggota kelas yang lebih dekat sehingga akurasi berkurang dan diusulkan metode klasifikasi lainnya
5. Salah satu metode *VMQDF* dengan klasifikasi *MQDF* memiliki pengenalan hingga 88%. Metode ini lebih baik untuk akurasi dibandingkan k-NN sehingga diteliti pengenalan *isolated character* dengan ekstraksi fitur penelitian (M. W. A. Kesiman et al., 2017) dan klasifikasi *MQDF* dan *DLQDF*

1.3 Batasan Penelitian

Sehubungan keterbatasan yang dimiliki, baik dari segi waktu, pemikiran serta biaya sehingga penelitian ini memiliki batas. Adapun ruang lingkup masalah terdiri dari:

1. Set data yang digunakan berasal dari tantangan 3 : *Isolated Character Recognition of Balinese Script in Palm Leaf Manuscript Images* salah satu bagian *AMADI_LontarSet*. Set data bisa diunduh di *ICFHR 2016 Challenge 3* . Data riset berupa data citra aksara Bali yang sudah tersegmentasi dengan baik. Hasil segmentasi kemudian diekstraksi fitur 3 metode dan digabung jadi 1. Hasil ekstraksi fitur digunakan untuk membuat kelas dan klasifikasi.
2. Penelitian ini menggunakan 4 ekstraksi fitur yaitu *HoG*, *NPWK*, *Zoning* dan gabungan 3 ekstraksi fitur
3. Penelitian ini meneliti klasifikasi *MQDF* dan *DLQDF*
4. Penelitian ini menggunakan seleksi fitur

1.4 Rumusan Masalah

Berdasarkan penjelasan latar belakang dapat dirumuskan permasalahan untuk ditemukan solusi sebagai berikut.

1. Bagaimana implementasi klasifikasi metode *MQDF* dan *DLQDF*
2. Bagaimana tingkat klasifikasi *MQDF* dan *DLQDF* terhadap ekstraksi fitur dan gabungan ekstraksi ?

1.5 Tujuan Penelitian

Adapun tujuan yang diharapkan dapat dicapai dari penelitian ini adalah sebagai berikut.

1. Mengetahui implementasi MQDF dan DLQDF
2. Mengetahui pengaruh klasifikasi dengan *MQDF* dan *DLQDF* terhadap ekstraksi fitur dan gabungan ekstraksi

1.6 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian guna memberi penjelasan kemanfaatan bagi pengembangan penelitian atau aplikasi.

Aspek Teoretis :

1. Menambah wawasan dan pengetahuan yang berhubungan dengan pengenalan aksara Bali pada lontar dan klasifikasi pada khususnya.
2. Peneliti dan pembaca dapat mengerti klasifikasi *MQDF* dan *DLQDF*.
3. Sebagai modal dasar untuk penelitian pengenalan aksara Bali di masa yang akan datang.
4. Secara teoretis diharapkan penelitian ini dapat meningkatkan dan mengembangkan penelitian dalam pengenalan aksara Bali. Penelitian menjadi tambahan dan penguatan teori yang sudah ada sebelumnya.

5. Harapan dalam penelitian ini dapat dijadikan masukan sebagai bahan pertimbangan dalam penyempurnaan hasil kajian lainnya.

Aspek Praktis

1. Diharapkan dalam penelitian ini dapat memberikan masukan kepada peneliti lainnya dalam memilih metode klasifikasi dalam pengenalan aksara.
2. Diharapkan dengan penelitian ini dapat membantu peneliti lainnya mengetahui cara kerja *MQDF* dan *DLQDF*.

