

Pembangkitan Dataset Aksara Bali Menggunakan Web Scrapping untuk Pengenalan Aksara Bali Berbasis Optical Character Recognition

I Gede Pradipta Adi Nugraha¹, Ahmad Asroni², Luh Joni Erawati Dewi³, Gede Indrawan⁴

Universitas Pendidikan Ganesha
Jl. Udayana 11, Singaraja, Buleleng, Bali, Indonesia

e-mail: pradiptaadinugraha97@gmail.com¹, ahmad.asroni@undiksha.ac.id²,
joni.erawati@undiksha.ac.id³, gindrawan@undiksha.ac.id⁴

Received : Agustus, 2023

Accepted : Agustus, 2023

Published : Agustus, 2023

Abstract

The dataset of extensive and comprehensive Balinese script images stands as a vital resource in the preservation of Bali's cultural heritage and the development of associated applications. This dataset enables in-depth analysis, character recognition, language processing, and application testing, including automatic script recognition, automated teaching, and language understanding, all contributing to the advancement of preserving Bali's cultural heritage. The research method employed to create this dataset involved accessing data sources, analyzing web structures, and utilizing web scraping techniques with JavaScript for automatic image retrieval. The two primary phases—initiation and data collection—facilitated the accumulation of a large dataset with high efficiency, expediting the data collection process and enhancing accuracy in Balinese script research. The dataset, sourced from Balinese, Indonesian, and English language dictionaries, comprised a total of 35,319 Balinese words, subsequently converted into Balinese script. The resultant dataset includes 35,319 pairs of Balinese script image texts and transliterations, holding significant potential for training Balinese script recognition models and language studies. This step solidifies the availability of a relevant, high-quality dataset critical in further technological advancements and continued research in Balinese language and script recognition.

Keywords: *balinese script, dataset, character recognition, web scraping, transliteration*

Abstrak

Dataset gambar aksara Bali yang besar dan komprehensif adalah sumber daya penting dalam upaya pemeliharaan warisan budaya Bali serta pengembangan aplikasi terkait. Dataset ini memungkinkan analisis mendalam, pengenalan karakter, pemrosesan bahasa, dan pengujian aplikasi seperti pengenalan tulisan otomatis, pengajaran otomatis, dan pemahaman bahasa, yang semuanya mendukung kemajuan dalam pelestarian warisan budaya Bali. Metode penelitian untuk menghasilkan dataset aksara Bali melibatkan akses ke sumber data, analisis struktur web, dan penerapan teknik web scraping dengan JavaScript untuk pengambilan gambar secara otomatis. Dua tahap utama, yaitu inisiasi dan pengumpulan data, memungkinkan pengumpulan dataset dalam jumlah besar dengan efisiensi tinggi, mempercepat proses pengumpulan data dan meningkatkan akurasi dalam penelitian aksara Bali. Data yang digunakan berasal dari kamus bahasa Bali, bahasa Indonesia, dan bahasa Inggris, dengan total 35.319 kata dalam bahasa Bali, yang kemudian dikonversi menjadi aksara Bali. Hasil dari pembuatan dataset ini mencakup 35.319 pasang data berupa gambar teks aksara Bali dan teks transliterasinya, memiliki potensi besar untuk pelatihan model pengenalan aksara Bali dan penelitian bahasa Bali. Langkah ini menguatkan ketersediaan dataset yang relevan, berkualitas tinggi, dan memiliki nilai signifikan dalam pengembangan teknologi serta penelitian lebih lanjut di bidang bahasa Bali dan pengenalan aksara Bali.

Kata Kunci: aksara bali, dataset, pengenalan karakter, web scraping, transliterasi

1. PENDAHULUAN

Aksara Bali merupakan sistem penulisan yang kaya dan kompleks yang digunakan dalam budaya Bali [1]. Berdasarkan aksara Brahmi, aksara Bali terdiri dari sekitar 50 simbol dasar yang dapat dikombinasikan dengan simbol vokal dan tanda baca khusus [2], [3]. Aturan penulisan dan tanda baca khusus dalam aksara Bali memengaruhi cara membaca dan mengucapkan kata-kata. Meskipun penggunaannya menurun dalam kehidupan sehari-hari, aksara Bali masih digunakan dalam konteks keagamaan dan seni budaya, serta dihargai sebagai warisan budaya penting yang sedang dilestarikan dan diajarkan kepada generasi muda di Bali [4]. Namun, mendapatkan dataset gambar aksara Bali dalam jumlah besar bisa menjadi tugas yang rumit dan memakan waktu jika dilakukan secara manual. Pengumpulan dataset yang besar berisi gambar-gambar aksara Bali merupakan salah satu proses penting dalam penelitian dan pengembangan aplikasi terkait [5]. Dengan memiliki dataset yang kaya dan luas, peneliti dan pengembang dapat melakukan analisis mendalam terhadap aksara Bali, termasuk pengenalan karakter, pemrosesan bahasa, dan pengembangan algoritma terkait. Dataset yang besar juga memungkinkan pengujian yang lebih akurat dan valid untuk mengembangkan aplikasi seperti pengenalan tulisan aksara Bali otomatis [6], pengajaran otomatis, atau pemahaman bahasa. Dengan memiliki sumber daya yang komprehensif seperti dataset gambar aksara Bali, kemajuan dalam pemeliharaan dan pelestarian warisan budaya Bali dapat diharapkan [7].

Mendapatkan dataset aksara Bali dalam jumlah besar menggunakan metode manual menghadapi tantangan yang signifikan. Proses pengumpulan manual memerlukan waktu, upaya, dan sumber daya manusia yang cukup besar [8]. Hal ini melibatkan pencarian, pengumpulan, dan pengolahan gambar aksara Bali dari berbagai sumber seperti buku, naskah, prasasti, dan artefak budaya. Tantangan lainnya adalah keberagaman gaya tulisan dan variasi aksara Bali dari waktu ke waktu, daerah ke daerah, dan individu ke individu. Oleh karena itu, memastikan keakuratan, konsistensi, dan representativitas dataset manual dapat menjadi tugas yang rumit [9]. Tantangan ini menekankan

pentingnya upaya untuk mengembangkan metode otomatis atau semi-otomatis yang dapat membantu dalam pengumpulan dataset aksara Bali dalam jumlah besar [10].

Salah satu metode pengumpulan dataset yang bisa menjadi rujukan untuk digunakan dalam mengumpulkan dataset aksara Bali adalah metode web scraping [11]. Web scraping adalah teknik pengambilan data otomatis dari halaman web yang dilakukan dengan menggunakan program komputer [12], [13]. Dalam proses ini, perangkat lunak yang disebut "scraper" digunakan untuk mengekstrak informasi yang relevan dari struktur halaman web. Scraper akan mengakses halaman web, mengidentifikasi elemen-elemen yang ditentukan seperti teks, gambar, tabel, atau hyperlink, dan mengekstrak data dari elemen tersebut [14]. Teknik ini memungkinkan pengumpulan data secara efisien dan otomatis tanpa perlu melakukan entri data secara manual. Web scraping telah digunakan dalam berbagai bidang, seperti riset, analisis pasar, pengumpulan berita, dan pengembangan aplikasi. Web scraping memiliki keunggulan dalam mengatasi keterbatasan pengumpulan manual dalam hal efisiensi, skalabilitas, dan aksesibilitas terhadap sumber daya digital yang luas.

Dalam hal efisiensi, web scraping memungkinkan pengambilan data secara otomatis dari halaman web dengan kecepatan yang jauh lebih tinggi dibandingkan metode manual. Dengan menggunakan teknik ini, pengumpulan data dapat dilakukan secara simultan dari berbagai sumber secara bersamaan [15]–[17]. Selain itu, web scraping juga memungkinkan skalabilitas, di mana jumlah data yang dapat diambil tidak terbatas oleh batasan fisik manusia. Dengan adanya aksesibilitas terhadap sumber daya digital yang luas, web scraping memungkinkan peneliti atau pengembang untuk mengumpulkan data dari berbagai situs web, platform media sosial, atau sumber daya digital lainnya tanpa terbatas pada batasan geografis atau fisik [18], [19]. Hal ini memungkinkan penelitian yang lebih komprehensif, analisis yang lebih mendalam, dan pengembangan aplikasi yang lebih canggih berdasarkan data yang luas dan bervariasi. Teknik web scraping efisien digunakan untuk mengumpulkan dataset aksara Bali dalam

jumlah besar dengan cepat. Dengan web scraping, scraper secara otomatis mengakses berbagai sumber digital yang berisi gambar aksara Bali, seperti situs perpustakaan digital, arsip online, atau koleksi digital. Scraper akan otomatis menavigasi halaman web, mengidentifikasi gambar aksara Bali, dan mengekstraknya secara massal [20], [21].

Dengan mengotomatiskan proses ini, web scraping memungkinkan pengumpulan data yang lebih cepat dan efisien dibandingkan dengan pencarian manual pada sumber fisik seperti buku atau prasasti [22]. Selain itu, web scraping juga memungkinkan pengumpulan data secara skalabilitas, di mana scraper dapat dijalankan pada berbagai sumber digital secara bersamaan, memungkinkan pengumpulan dataset aksara Bali dalam jumlah besar dalam waktu yang lebih singkat [23], [24]. Langkah-langkah teknis dalam menggunakan web scraping untuk mengumpulkan gambar aksara Bali meliputi beberapa tahapan yaitu:

- a) Mengakses sumber website yang berisi gambar aksara Bali.
- b) Menganalisis struktur halaman web untuk mengidentifikasi elemen yang berisi gambar aksara Bali pada halaman website tersebut.
- c) Memilih alat web scraping yang sesuai seperti Beautiful Soup atau Scrapy.
- d) Menggunakan alat web scraping untuk mengakses halaman web secara otomatis dan mengekstrak gambar aksara Bali.
- e) Menyimpan dan mengolah gambar yang berhasil diambil.
- f) Mengulangi langkah-langkah di atas untuk mengumpulkan semua gambar aksara Bali dari sumber halaman website.

Penggunaan dataset aksara Bali yang dihasilkan melalui web scraping memiliki potensi yang luas dalam berbagai konteks. Pertama, dataset ini dapat digunakan dalam pengenalan karakter aksara Bali secara otomatis. Dengan menggunakan teknik pembelajaran mesin atau pengenalan pola, dataset dapat membantu melatih model yang mampu mengenali dan mengklasifikasikan karakter aksara Bali dengan akurasi tinggi [25]. Hal ini dapat digunakan dalam pengembangan aplikasi pengenalan tulisan aksara Bali otomatis atau alat bantu penulisan. Selain itu, dataset aksara Bali yang luas dapat digunakan dalam analisis linguistik dan pemahaman bahasa Bali [26]. Data ini dapat

digunakan untuk menganalisis struktur kalimat, pola kata, atau perubahan bahasa dalam teks berbahasa Bali. Analisis ini dapat memberikan wawasan tentang sejarah, struktur, dan karakteristik bahasa Bali yang lebih mendalam. Dataset ini juga dapat digunakan dalam pengembangan aplikasi atau sistem pemrosesan bahasa alami untuk bahasa Bali. Dataset aksara Bali juga sangat berharga dalam penelitian budaya Bali. Data ini dapat digunakan untuk mempelajari teks-teks kuno, naskah-naskah suci, atau cerita-cerita tradisional yang ditulis dalam aksara Bali [27]. Penelitian ini dapat memberikan pemahaman yang lebih baik tentang kebudayaan Bali, tradisi, dan warisan sastra Bali. Dataset juga dapat digunakan dalam digitalisasi dan pelestarian naskah-naskah yang langka dan rapuh, sehingga memungkinkan akses yang lebih luas terhadap warisan budaya Bali yang kaya.

Penelitian ini memiliki nilai dan keunikan yang signifikan dalam konteks penggunaan web scraping untuk mengumpulkan dataset aksara Bali. Dengan menggunakan teknik web scraping, penelitian ini mampu mengatasi keterbatasan pengumpulan manual yang memakan waktu dan sulit dalam mengumpulkan gambar aksara Bali dalam jumlah besar [28]. Dalam prosesnya, penelitian ini dapat memanfaatkan sumber daya digital yang website yang didalamnya terkandung sumber daya aksara Bali yang dapat diekstraksi sesuai dengan kebutuhan, yang memberikan aksesibilitas yang lebih tinggi terhadap sumber data. Dengan menggunakan web scraping, penelitian ini memungkinkan pengumpulan data secara cepat, efisien, dan skalabilitas yang besar [29], sehingga memungkinkan analisis yang lebih mendalam tentang aksara Bali dalam berbagai konteks, termasuk pengenalan karakter, analisis linguistik, pengembangan aplikasi berbasis aksara Bali, dan penelitian budaya Bali [30].

2. METODE PENELITIAN

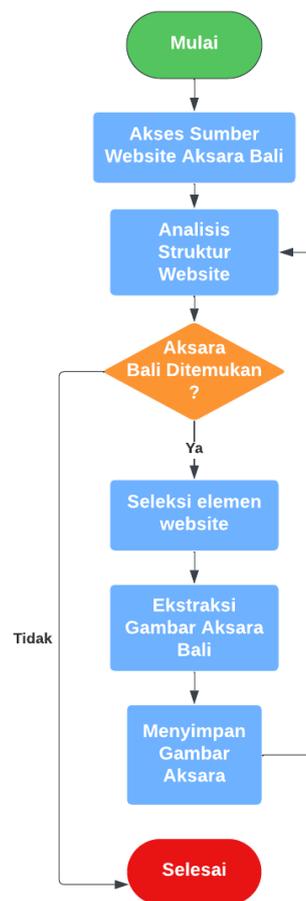
Metode penelitian yang digunakan dalam menghasilkan dataset aksara Bali melibatkan serangkaian tahapan. Pertama, dilakukan akses ke dalam website sumber data aksara Bali. Kemudian, dilakukan eksplorasi terhadap sumber website yang berisi gambar aksara Bali, diikuti dengan analisis struktur halaman web untuk mengidentifikasi elemen yang berisi gambar aksara Bali. Langkah berikutnya adalah

melakukan seleksi elemen HTML dan menggunakan teknik web scraping dengan JavaScript untuk mengambil gambar secara otomatis dari halaman web tersebut. Data gambar yang berhasil diambil selanjutnya disimpan dan diolah guna analisis lebih lanjut atau keperluan penggunaan berikutnya.

Secara garis besar, tahapan di atas dapat dikelompokkan menjadi dua fase: fase inisiasi dan fase pengumpulan data. Tahap inisiasi melibatkan akses ke sumber data dan eksplorasi awal, sedangkan tahap pengumpulan data melibatkan proses pengambilan gambar secara

otomatis menggunakan teknik web scraping. Metode ini memberikan kemampuan kepada peneliti untuk mengumpulkan gambar aksara Bali dalam jumlah besar secara efektif dan otomatis, yang pada gilirannya mempercepat proses pengumpulan data dan meningkatkan efisiensi penelitian.

Tahapan-tahapan dalam metode penelitian ini dan cara kerja dari web scraping dalam menghasilkan dataset aksara Bali dapat dilihat secara lebih jelas dalam Gambar 1. Metode Penelitian.



Gambar 1. Metode Penelitian

2.1 Tahapan Inisiasi

Pada tahapan awal penelitian, analisis dilakukan untuk mengidentifikasi jenis data yang akan diambil, memfasilitasi proses pengumpulan data dari situs web yang bersangkutan. Tahap ini menggunakan bahasa pemrograman JavaScript, sering dipakai dalam pengembangan aplikasi berbasis web. Awalnya, JavaScript

dikembangkan untuk meningkatkan interaktivitas dan dinamika aplikasi web melalui manipulasi elemen HTML. Popularitas JavaScript sebagai bahasa utama untuk web scraping disokong oleh sejumlah keunggulan yang signifikan. Halaman web sumber yang akan di-scraping ditampilkan pada Gambar 2. Website Sumber Aksara Bali.


```

1  $(document).ready(function() {
2      async function f() {
3          let promise = new Promise((resolve, reject) => {
4              html2canvas(
5                  document.getElementById("html-content-holder"),
6                  {
7                      allowTaint: true,
8                      useCORS: true
9                  }
10             ).then(function (canvas) {
11                 var anchorTag = document.createElement("a");
12                 anchorTag.download = "aksara_{{ $a }}.png";
13                 anchorTag.href = canvas.toDataURL();
14                 anchorTag.target = '_blank';
15                 anchorTag.click();
16             });
17             setTimeout(() => resolve("done!"), 3000)
18         });
19
20         let result = await promise; // wait until the promise resolves (*)
21     }
22
23     f().then(function() {
24         if ('{{ $a }}' != '{{ $b }}') {
25             var a = parseInt('{{ $a }}') + 1;
26
27             window.location.href = 'url('generatedataset/')' + a + '/' + '{{ $b }}';
28         } else {
29             alert('{{ $a }}');
30         }
31     });
32 });

```

Gambar 4. Cuplikan Kode Ekstraksi Data

```

1  let jumdata = '{{ $data->count() }}';
2
3  for (let index = 0; index < jumdata; index++) {
4      var text = $('#text' + index).val();
5      saveStaticDataToFile((index + 1), text);
6  }
7
8  function saveStaticDataToFile(index, text) {
9      var blob = new Blob([text], {
10         type: "text/plain;charset=utf-8"
11     });
12
13     // saveAs(blob, "aksara_" + (index) + ".gt.txt");
14 }

```

Gambar 5. Cuplikan Kode Simpan Hasil Ekstraksi Gambar dan Gound Truth Aksara Bali

2.4 Tahapan Menyimpan Data

Tahapan ini berfokus pada proses penyimpanan data yang telah melewati proses ekstraksi. Langkah ini membuat transformasi data ke dalam format yang sesuai dengan sifat masing-masing jenis data, serta memastikan kesiapan data tersebut untuk langkah analisis yang lebih dalam.

Pada tahap awal, langkah kunci adalah memilih format penyimpanan yang paling sesuai dengan karakteristik unik dari tiap jenis data. Ketika datang ke gambar aksara Bali, format .png terpilih karena mampu mempertahankan kualitas visual yang esensial. Sementara itu, data transliterasi ditempatkan dalam format .txt yang sederhana namun efektif, mampu menyimpan teks dengan kejelasan struktur. Keputusan ini membawa dampak signifikan dalam memastikan data dapat diakses, dikelola, dan digunakan secara efisien untuk keperluan analisis di masa yang akan datang.

Setelah format penyimpanan ditetapkan, tahapan berikutnya adalah menempatkan data dalam format yang telah disesuaikan tersebut. Gambar aksara Bali dikemas dalam format .png, mempertahankan kemiripan visual yang penting bagi data semacam ini. Sementara itu, data transliterasi, yang diwakili dalam bentuk teks, diorganisir dalam format .txt yang simpel namun efektif. Proses ini melibatkan pengaturan data dalam struktur direktori yang teratur sesuai jenis dan formatnya. Dengan pengorganisasian yang terstruktur ini, data siap untuk dipergunakan dan dikelola dengan efektivitas yang diperlukan dalam langkah analisis yang akan datang.

Melalui tahapan ini, penyimpanan data yang terstruktur dan format yang sesuai untuk tiap jenis data memainkan peran penting dalam memastikan data yang telah melewati proses ekstraksi dapat terus diakses, dikelola, dan digunakan dengan kemudahan. Pendekatan ini mengarah pada kesiapan data untuk langkah

analisis berikutnya, sambil menjunjung tinggi keberlanjutan dan efisiensi dalam pengelolaan data.

3. HASIL DAN PEMBAHASAN

Bagian ini membahas hasil penelitian serta proses ekstraksi data terkait aksara Bali, meliputi dua dataset utama: dataset gambar aksara Bali dan dataset teks transliterasi. Dataset gambar aksara Bali berisi gambar-gambar aksara yang berhasil diekstraksi dari sumber web, disimpan dalam format .png, sementara dataset teks transliterasi yang memuat transliterasi teks aksara Bali menjadi format .txt. Melalui tahap seleksi elemen, elemen-elemen yang mengandung informasi aksara Bali diidentifikasi pada halaman web, dan data diekstraksi dari elemen-elemen tersebut. Kedua dataset ini akan digunakan untuk menganalisis bentuk, variasi, dan karakteristik visual aksara, serta untuk memahami arti dan konteks dari setiap aksara Bali.

3.1 Komposisi Hasil Dataset Aksara Bali

Hasil ekstraksi dataset dibuat melalui proses web scraping dengan tujuan utama menghasilkan dua jenis dataset yaitu dataset gambar aksara Bali dan ground truth transliteration (transliterasi sebenarnya). Proses yang dilalui untuk menghaikan dua jenis dataset ini melibatkan konversi data bahasa Bali ke dalam aksara Bali menggunakan Unicode. Dataset bahasa Bali yang digunakan dalam penelitian ini adalah dataset transliterasi bahasa Bali yang terdiri dari 35.319 kata.

Komposisi dataset transliterasi ini mencakup kata-kata dalam bahasa Bali, bahasa Indonesia, dan bahasa Inggris. Jumlah data yang terdapat dalam dataset ini dapat dilihat lebih jelas dalam Tabel 1: Komposisi Dataset Transliterasi. Tabel tersebut memberikan gambaran yang jelas tentang distribusi data berdasarkan indeks kata dalam dataset transliterasi yang telah digunakan yang berhasil diekstraksi menjadi dataset Aksara Bali.

Tabel 1: Komposisi Dataset Transliterasi

No.	Indeks Huruf Latin	Indeks Aksara Bali	Jumlah Dataset Image & Text Transliterasi
1	A	ꦱꦏꦶ	1423
2	B	ꦁꦏꦶ	2090

No.	Indeks Huruf Latin	Indeks Aksara Bali	Jumlah Dataset Image & Text Transliterasi
3	C	ꦑꦶ	1171
4	D	ꦕꦶ	936
5	E	ꦒ	642
6	G	ꦒꦶ	1686
7	H	ꦒꦶ	28
8	I	ꦒꦶ	468
9	J	ꦒꦶ	792
10	K	ꦒꦶ	3767
11	L	ꦒꦶ	1494
12	M	ꦒꦶ	4881
13	N	ꦒꦶ	4602
14	O	ꦒꦶ	274
15	P	ꦒꦶ	3508
16	R	ꦒꦶ	894
17	S	ꦒꦶ	2943
18	T	ꦒꦶ	2279
19	U	ꦒꦶ	856
20	W	ꦒꦶ	576
21	Y	ꦒꦶ	9

3.2 Hasil Ekstraksi Dataset Aksara Bali

Dari komposisi dataset transliterasi berhasil diperoleh dua jenis dataset, yaitu dataset gambar dan dataset teks transliterasi yang berpasangan. Dataset teks transliterasi ini menjadi acuan utama (ground truth) yang sesuai dengan setiap gambar aksara Bali yang berhasil dihasilkan. Berdasarkan data komposisi dan jumlah indeks kata dataset yang tertera dalam tabel di atas, maka dalam proses generate datasetnya berhasil menghasilkan total 34.187 dataset aksara Bali berpasangan dengan ground

truth transliterasi dalam bentuk file teks dari komposisi tersebut. Pada tahapan ekstraksi ini juga dilakukan proses pengecekan dan proses kontrol ketika akan menyimpan data hasil ekstraksi, kedua format dataset akan disimpan ketika sudah berhasil diekstraksi, sehingga tidak ada dataset yang dihasilkan dari proses ini tidak memiliki pasangan. Berikut ini pada Gambar 6. Contoh Hasil Generate Dataset Gambar dan Gound Truth Aksara Bali adalah cuplikan contoh hasil dataset yang berhasil dihasilkan melalui proses generate dataset menggunakan metode web scrapping.



Gambar 6. Contoh Hasil Generate Dataset Gambar dan Gound Truth Aksara Bali

4. KESIMPULAN

Pembuatan dataset Aksara Bali melalui metode web scraping melibatkan beberapa tahap utama. Tahap inisiasi meliputi persiapan transliterasi bahasa Bali dan penggunaan Unicode untuk mengonversi huruf Latin Bali ke aksara Bali. Selanjutnya, dilakukan seleksi elemen HTML yang digunakan sebagai sumber dataset, kemudian dilakukan ekstraksi data yang berupa gambar aksara Bali beserta transliterasinya. Sebelum disimpan, dilakukan pemeriksaan agar pasangan data terjamin lengkap. Proses ini krusial karena memastikan ketersediaan dataset yang relevan dan berkualitas tinggi untuk penelitian berikutnya, seperti pelatihan model pengenalan aksara Bali atau penelitian bahasa Bali.

Data yang diambil dari kamus bahasa Bali, Indonesia, dan Inggris berjumlah 35.319 kata dalam bahasa Bali. Data ini dikonversi ke aksara Bali, menghasilkan 35.319 gambar aksara Bali dan 35.319 ground truth transliterasi dalam format .png dan .gt.txt. Dataset ini memiliki nilai penting dalam pengembangan teknologi atau penelitian lebih lanjut dalam bahasa Bali dan pengenalan aksara Bali. Data ini kemudian mengalami proses konversi menjadi bentuk aksara Bali. Selanjutnya aksara bali tersebut diproses generasi dataset dan menghasilkan sepasang data set, yang terdiri dari gambar teks satu baris dengan ekstensi file .png dan transliterasi satu baris dengan ekstensi file .gt.txt. Keseluruhan jumlah set data yang dihasilkan dari proses ini adalah sebanyak 35.319 gambar aksara Bali dan 35.319 ground truth transliterasinya. Dataset ini memiliki

potensi besar dan dapat digunakan dalam berbagai konteks, seperti pelatihan model pengenalan aksara Bali atau dalam penelitian bahasa Bali. Keseluruhan proses ini merupakan langkah kunci dalam penelitian, dan mengukuhkan ketersediaan data yang relevan, berkualitas tinggi, dan memiliki nilai signifikan untuk pengembangan teknologi atau penelitian lebih lanjut di bidang bahasa Bali dan pengenalan aksara Bali.

PERNYATAAN PENGHARGAAN

Penelitian di bidang teknologi untuk pengumpulan dan pengelolaan data terkait kearifan lokal di Indonesia ini mendapat dukungan dari Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

DAFTAR PUSTAKA

- [1] G. Indrawan, I. K. Paramarta, and K. Agustini, "A New Method of Latin-to-Balinese Script Transliteration Based on Noto Sans Balinese Font and Dictionary Data Structure," in Proceedings of the 2nd International Conference on Software Engineering and Information Management, in ICSIM '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 75–79. doi: 10.1145/3305160.3305167.
- [2] G. Indrawan, Sariyasa, and I. K. Paramarta, "A New Method of Latin-To-Balinese Script Transliteration based on Bali Simbar Font," in 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019, pp. 1–6.

- doi: 10.1109/ICIC47613.2019.8985675.
- [3] G. Indrawan, I. K. Paramarta, and K. Agustini, "A new method of Latin-to-balinese script transliteration based on noto sans balinese font and dictionary data structure," in *ACM Int. Conf. Proceeding Ser.*, 2019, pp. 75–79.
- [4] A. Asroni, G. Indrawan, and L. Joni Erawati Dewi, "Implementasi Hirarki Dataset Dalam Membangun Model Language Aksara Bali Menggunakan Framework Tesseract OCR", [Online]. Available: <https://s.id/jurnalresistor>
- [5] G. Indrawan, I. Gede Nurhayata, and I. Ketut Paramarta, "A Method to Accomodate Backward Compatibility on the Learning Application-based Transliteration to the Balinese Script." [Online]. Available: www.ijacsa.thesai.org
- [6] L. H. Loekito, G. Indrawan, and I. K. Paramarta, "Error Analysis of Latin-to-Balinese Script Transliteration Method Based on Noto Sans Balinese Font," 2020.
- [7] G. Indrawan, I. P. E Swastika, and I. K. Paramarta, "An Improved Algorithm and Accuracy Analysis Testing Cases of Latin-to-Balinese Script Transliteration Method based on Bali Simbar Dwijendra Font".
- [8] G. Indrawan, K. Setemen, W. Sutaya, and I. K. Paramarta, "Handling of Line Breaking on Latin-to-Balinese Script Transliteration Web Application as Part of Balinese Language Ubiquitous Learning," in *2020 6th International Conference on Science in Information Technology (ICSITech)*, 2020, pp. 40–44. doi: 10.1109/ICSITech49800.2020.9392035.
- [9] G. Indrawan, I. K. Paramarta, K. Agustini, and Sariyasa, "Latin-to-Balinese script transliteration method on mobile application: A comparison," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 3, pp. 1331–1342, 2018, doi: 10.11591/ijeecs.v10.i3.pp1331-1342.
- [10] G. Indrawan, I. K. Paramarta, K. Agustini, and Sariyasa, "Latin-to-Balinese script transliteration method on mobile application: A comparison," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 10, no. 3, pp. 1331–1342, 2018.
- [11] G. Indrawan, N. N. H. Puspita, I. K. Paramarta, and Sariyasa, "LBtrans-bot: A Latin-to-Balinese script transliteration robotic system based on noto sans Balinese font," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 1247–1256, Dec. 2018, doi: 10.11591/ijeecs.v12.i3.pp1247-1256.
- [12] G. Indrawan, N. N. H. Puspita, I. K. Paramarta, and Sariyasa, "LBtrans-bot: A Latin-to-Balinese script transliteration robotic system based on noto sans Balinese font," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 1247–1256, 2018, doi: 10.11591/ijeecs.v12.i3.pp1247-1256.
- [13] I. K. Paramarta, G. Indrawan, I. B. Rai, and I. N. Martha, "Bound Vowels Grapheme Representation in Balinese Script," in *Proceedings of the 2nd International Conference on Languages and Arts across Cultures (ICLAAC 2022)*, Atlantis Press SARL, 2023, pp. 165–172. doi: 10.2991/978-2-494069-29-9_18.
- [14] G. Indrawan, I. G. Aris Gunadi, M. Santo Gitakarma, and I. K. Paramarta, "Latin to Balinese Script Transliteration: Lessons Learned from the Computer-Based Implementation," in *Proceedings of the 2021 4th International Conference on Software Engineering and Information Management*, in *ICSIM '21*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 171–175. doi: 10.1145/3451471.3451499.
- [15] G. Indrawan, G. Rasben Dantes, K. Yota Ernanda Aryanto, and I. Ketut Paramarta, "Handling of Mathematical Expression on Latin-to-Balinese Script Transliteration Method on Mobile Computing," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, Nov. 2020, pp. 1–5. doi: 10.1109/ICIC50835.2020.9288563.
- [16] G. Indrawan, A. Asroni, L. Joni Erawati Dewi, I. G. A. Gunadi, and I. K. Paramarta, "Balinese Script Recognition Using Tesseract Mobile Framework," *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 13, no. 3, p. 160, Nov. 2022, doi: 10.24843/lkjiti.2022.v13.i03.p03.
- [17] M. R. Rafsanjani, "ScrapPaper: A web

- scrapping method to extract journal information from PubMed and Google Scholar search result using Python," bioRxiv, 2022, [Online]. Available: <https://api.semanticscholar.org/CorpusID:247412163>
- [18] G. Indrawan, L. J. E. Dewi, I. Gede Aris Gunadi, K. Agustini, and I. Ketut Paramarta, "The Analysis of Noto Serif Balinese Font to Support Computer-assisted Transliteration to Balinese Script," in *Information and Communication Technology for Competitive Strategies (ICTCS 2021)*, A. Joshi, M. Mahmud, and R. G. Ragel, Eds., Singapore: Springer Nature Singapore, 2023, pp. 571–580.
- [19] G. Adomavicius and A. Tuzhilin, "Web Scraping: State of the art," *IEEE Trans Knowl Data Eng*, vol. 17, no. 6, pp. 734–749, 2019.
- [20] G. Indrawan, A. Asroni, L. Joni Erawati Dewi, I. G. A. Gunadi, and I. K. Paramarta, "Balinese Script Recognition Using Tesseract Mobile Framework," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 13, no. 3, p. 160, Nov. 2022, doi: 10.24843/lkjiti.2022.v13.i03.p03.
- [21] S. Chaudhari, R. Aparna, V. G. Tekkur, G. L. Pavan, and S. R. Karki, "Ingredient/Recipe Algorithm using Web Mining and Web Scraping for Smart Chef," *Proceedings of CONECCT 2020 - 6th IEEE International Conference on Electronics, Computing and Communication Technologies*, no. 3, pp. 22–25, 2020, doi: 10.1109/CONECCT50063.2020.9198450.
- [22] G. Indrawan, I. W. Sutaya, K. U. Ariawan, M. S. Gitakarma, I. G. Nurhayata, and I. K. Paramarta, "A METHOD FOR NON-ALPHANUMERIC TEXT PROCESSING ON TRANSLITERATION TO THE BALINESE SCRIPT," *ICIC Express Letters*, vol. 16, no. 7, pp. 687–694, Jul. 2022, doi: 10.24507/icicel.16.07.687.
- [23] G. Indrawan, C. O. Birawidya, L. J. Erawati Dewi, K. Agustini, I. Gede Aris Gunadi, and I. Ketut Paramarta, "DERIVATIVE WORD CONVERSION METHOD TO BALINESE SCRIPT ON MOBILE COMPUTING," *ICIC Express Letters*, vol. 17, no. 7, pp. 725–733, Jul. 2023, doi: 10.24507/icicel.17.07.725.
- [24] M. J. Lee, J. Kang, K. Hreha, and M. Pappadis, "A Novel Web Scraping Approach to Identify Stroke Outcome Measures: A Feasibility Study," *Arch Phys Med Rehabil*, vol. 103, no. 3, p. e30, 2022, doi: <https://doi.org/10.1016/j.apmr.2022.01.082>.
- [25] V. A. Flores, P. A. Permatasari, and L. Jasa, "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword," *Majalah Ilmiah Teknologi Elektro*, vol. 19, no. 2, p. 157, 2020, doi: 10.24843/mite.2020.v19i02.p06.
- [26] I. M. D. R. Mudiarta et al., "Balinese character recognition on mobile application based on tesseract open source OCR engine," *J Phys Conf Ser*, vol. 1516, no. 1, 2020, doi: 10.1088/1742-6596/1516/1/012017.
- [27] P. Nyoman Crisnapati et al., "Pasang Aksara Bot: A Balinese Script Writing Robot using Finite State Automata Transliteration Method," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2019, doi: 10.1088/1742-6596/1175/1/012108.
- [28] A. S. -, R. K. -, K. P. -, M. Kr. R. -, and V. S. -, "E-commerce Price Comparison Website using Web Scraping," *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, vol. 11, no. 3, Jun. 2023, doi: 10.37082/ijirms.v11.i3.230223.
- [29] A. Ahmed and A. Salam, "Automatic Scientific Literature Gathering and Analysis from Textual Corpus using Web Scraping and Locality Sensitive Hashing," 2023, doi: 10.0825/IEEESEM.2023156577.
- [30] W. Uriawan, A. Wahana, D. Wulandari, W. Darmalaksana, and R. Anwar, "Pearson correlation method and web scraping for analysis of islamic content on instagram videos," *Proceedings - 2020 6th International Conference on Wireless and Telematics, ICWT 2020, 2020*, doi: 10.1109/ICWT50448.2020.9243626.