

## Pengenalan Aksara Kaganga Lampung dengan Menggunakan Metode K-Nearest Neighbour (K-NN)

Swadexi Istiqphara<sup>1\*</sup>, Afida Nur Faida<sup>2</sup>, Anisa Ulya Darajat<sup>3</sup>

Teknik Elektro - Institut Teknologi Sumatera, Lampung<sup>1</sup>

Dinas Komunikasi dan Informatika, Kabupaten Tulang Bawang Lampung<sup>2</sup>

Teknik Elektro, Universitas Lampung, Lampung<sup>3</sup>

\*swadexi.istiqphara@el.itera.ac.id<sup>1</sup>

### ABSTRACT

*Character recognition technology has been carried out using various artificial intelligence methods. In this paper, one of the artificial intelligence methods, namely K-nearest Neighbor (K-NN), will be proposed to recognize the kaganga script, which is the Lampung province script. The Lampung kaganga script is one of the Indonesian cultural assets from the native Lampung tribe. Currently, only a few people understand this Lampung kaganga script. Therefore, an effort is needed to preserve this character, one of which is through this research so that a computer application can be produced that helps in learning the Lampung kaganga script. In this paper, the recognition of the Kaganga Lampung script using the K-Nearest Neighbor (K-NN) will measure performance in terms of the recognition processing time and the recognition accuracy of this Lampung kaganga script. From the tests that have been carried out, the results obtained by using the K-NN method are able to recognize the kaganga Lampung character with an accuracy of 70% and the average time of the character recognition process is 1.1412 seconds.*

**Keywords :** Artificial Intelligence, character recognition, K-Nearest Neighbour, Kaganga, Lampung Characters

### INTISARI

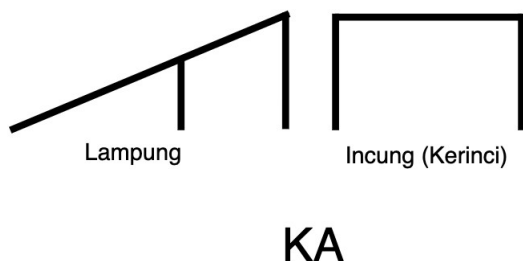
Teknologi pengenalan karakter telah banyak dilakukan dengan menggunakan berbagai metode kecerdasan buatan. Dalam paper ini salah satu metode kecerdasan buatan yaitu K-nearest Neighbour (K-NN) akan diusulkan untuk mengenali aksara kaganga yang merupakan aksara provinsi Lampung. Aksara kaganga lampung merupakan salah satu aset budaya indonesia dari suku asli lampung. Saat ini hanya sedikit masyarakat yang memahami aksara kaganga lampung ini. Oleh karena itu diperlukan suatu usaha untuk melestarikan karakter ini salah satunya yaitu melalui penelitian ini agar dapat dihasilkannya aplikasi komputer yang membantu dalam mempelajari aksara kaganga lampung. Pada paper ini, rekognisi aksara Kaganga Lampung dengan menggunakan K-Nearest Neighbour(K-NN) akan diukur performansi dalam hal waktu proses rekognisi dan keakurasian pengenalan aksara kaganga lampung ini. Dari pengujian yang telah dilakukan, diperoleh hasil dengan menggunakan metode K-NN mampu untuk mengenali karakter kaganga lampung dengan akurasi sebesar 70% dan waktu rata-rata proses rekognisi aksara sebesar 1.1412 second.

Kata kunci: Kecerdasan buatan, pengenalan karakter, K-Nearest Neighbour, Kaganga, Aksara Lampung

### I. PENDAHULUAN

Indonesia memiliki banyak suku budaya yang tersebar di beberapa provinsi di Indonesia [1][2]. Dalam satu provinsi di Indonesia dapat terdiri dari lebih satu suku, dimana setiap suku tersebut dapat memiliki bahasa dan aksara sendiri. Aksara kaganga merupakan aksara daerah yang tersebar di beberapa provinsi di pulau sumatera, meskipun memiliki nama yang sama, aksara Kaganga

memiliki bentuk karakter yang berbeda antara satu daerah dengan daerah lainnya meskipun memiliki kemiripan seperti yang ditunjukkan pada Gambar 1 yang merupakan huruf Ka yang ditulis dalam aksara lampung [3][4] dan akasara incung dari Kerinci[5][6].



**Gambar 1.** Perbandingan Aksara Kaganga Kerinci dan Lampung

Saat ini, pembelajaran terkait aksara daerah masih sedikit [7] dan diperlukan suatu inovasi agar pelajaran terkait aksara daerah menarik untuk dipelajari baik bagi anak-anak maupun dewasa. Salah satu upaya yang dapat dilakukan pada era ini adalah dengan membuat aplikasi yang mampu membawa pengguna untuk berperan aktif, sebagai contoh aplikasi tersebut mampu memeriksa apakah aksara yang ditulis oleh pengguna sudah sesuai atau belum. Untuk mencapai hal tersebut, diperlukan penelitian terkait sistem pengenalan citra aksara yang akan dilakukan pada penelitian ini.

Beberapa penelitian tentang rekognisi karakter sudah banyak dilakukan, di antaranya adalah dengan menggunakan metode *fully convolutional neural networks* [8], *Machine Learning Classifier* [9], Support Vector Machine Classifier [10], Teknik Shape decomposition [11], dan K-Nearest Neighbour (K-NN) [12]. Pada penelitian ini metode K-NN akan digunakan untuk mengenali aksara kaganga lampung ini. Metode K-NN dipilih karena kemudahan dalam menggunakannya, sehingga melalui penelitian ini, diharapkan dapat menjadi referensi terkait teknik dan algoritma pengenalan karakter kaganga ini.

## II. METODE PENELITIAN

Pada penelitian ini aksara Kaganga Lampung yang akan dikenali terdiri dari 20 karakter seperti yang ditunjukkan pada Gambar 2.

KA	GA	NGA	PA	BA
MA	TA	DA	NA	CA
JA	NYA	YA	A	LA
RA	SA	WA	HA	GHA

**Gambar 2.** Aksara Kaganga Lampung

Terdapat dua tahap yang dilakukan untuk mengenali aksara kaganga lampung, yaitu :

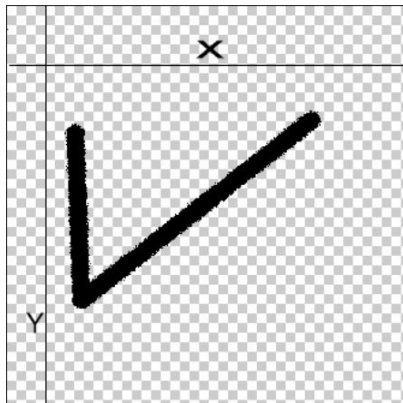
1. *Tahap Pelatihan*, merupakan tahap memperoleh data-data dari contoh gambar yang diberikan, sistem dapat dilatih dengan menggunakan karakter yang sama tapi dengan pola yang sedikit berbeda, hal ini bertujuan untuk memberitahukan kepada sistem bahwa terdapat perbedaan pola meskipun karakter sama.
2. *Tahap Pengujian*, merupakan tahap dimana sistem diuji dengan diberikan beberapa karakter yang ditulis sedikit berbeda dengan yang diberikan pada saat tahap pelatihan.

Pada metode *K-Nearest Neighbour* (K-NN) menggunakan proses kerja yang sama dengan metode kecerdasan buatan lainnya, yaitu dengan melakukan tahap pelatihan (*training*) dan tahap pengujian (*testing*) menguji datanya.

### A. Tahap Pelatihan

Adapun proses yang dilakukan pada tahap pelatihan adalah sebagai berikut:

- 1) Ekstraksi karakter digunakan untuk mengetahui nomor pixel karakter awal dan akhir. Seperti yang ditunjukkan pada Gambar 3, dimana pada gambar ini akan kita peroleh pixel pertama sumbu ( $Px_{min}$ ) adalah pada pixel ke 7 dan diakhiri pada pixel  $Px_{max} = 33$ , sedangkan pada sumbu Y ( $Py_{min}$ ) = 13 dan ( $Py_{max}$ ) = 32.



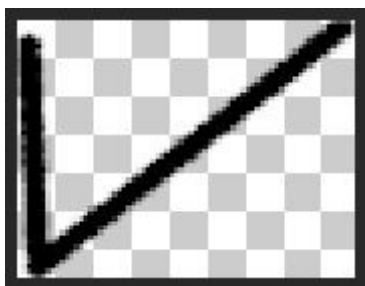
Gambar 3. Posisi karakter pada gambar

- 2) Normalisasi karakter agar memiliki ukuran pixel yang sama yaitu dengan formulasi sebagai berikut:

$$SkalaX = 9 / (Px_{max} - Px_{min}) * pixel\_x\_character \quad (1)$$

$$SkalaY = 9 / (Py_{max} - Py_{min}) * pixel\_y\_character \quad (2)$$

Tujuan dari normalisasi ini adalah supaya setiap gambar dengan resolusi yang berbeda-beda memiliki resolusi yang sama yaitu maksimum 9 pixel x dan 9 pixel y. Hasil normalisasi gambar 1 akan menjadi seperti pada Gambar 4.



Gambar 4. Hasil Normalisasi

Dari gambar 4 dapat dilihat hasil proses normalisasi gambar menunjukkan bahwa pada sumbu x karakter hanya terdiri dari 9 Pixel, sedangkan pada sumbu y hanya 7 pixel.

- 1) Penggabungan data, pada tahap ini adalah menggabungkan titik pixel dengan tujuan untuk mengumpulkan informasi koordinat pixel pada sumbu x dan y menjadi satu data, dimana pada sumbu X merupakan nilai puluhan sedangkan pada sumbu y merupakan nilai satuan. Sehingga diperoleh formulasi nya adalah sebagai berikut :

$$data(k) = \sum_{i=1}^{Px_{max}} \sum_{j=1}^{Py_{max}} (pixelX(i) * 10 + pixelY(j)) \quad (3)$$

- 2) Setelah data diperoleh maka data tersebut akan menjadi *knowledge data* atau pengetahuan yang dapat disimpan kedalam sebuah berkas basis data. Beberapa basis data pengetahuan dapat disimpan sebuah *file* saja.

#### B. Tahap Pengujian

Tahap ini merupakan tahap dimana dilakukannya proses pencocokan dengan basis data pengetahuan yang sudah dilakukan pada tahap 1. Langkah – langkah yang dilakukan memiliki langkah yang sama dalam tahap pelatihan yaitu langkah 1-3, langkah ini dilakukan untuk menyiapkan data yang akan di uji kecocokan dengan basis data pengetahuan. Kemudian langkah selanjutnya adalah:

1. Mencocokkan dengan basis data pengetahuan dengan cara menggunakan perhitungan *root mean square error (RMSE)*. Seperti yang ditunjukkan pada Persamaan (4)

$$m = \sqrt{(data\_uji - data\_p(k))^2} \quad (4)$$

2. Mencari nilai terkecil dari nilai *m*, dan nilai indeks *k* yang memiliki nilai *m* terkecil

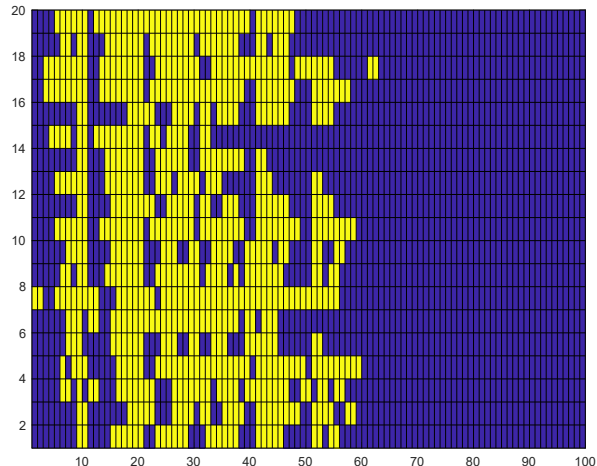
$$m\_minimum = \min(m) \\ k = \text{find}(m == m\_minimum)$$

3. Indeks *k* merupakan indeks basis data yang memiliki nilai paling dekat dengan data uji, sehingga indeks *k* adalah solusi dari hasil uji sistem.

### III. HASIL PENELITIAN DAN PEMBAHASAN

Pengujian dilakukan dengan melatih sistem pengenalan aksara dengan memberikan dua puluh

karakter konsonan. Tahap pelatihan dilakukan dengan memberikan beberapa sampel karakter dari masing-masing aksara. Sampel tersebut dapat berupa karakter yang sama tapi memiliki sedikit perbedaan pada penulisan aksara.



**Gambar 5.** Basis Data Pengetahuan

Pada proses training, semakin banyak sampel tiap aksara maka akan semakin baik dalam pengenalan karakter yang digunakan, hal ini akan memudahkan sistem dalam mengenali berbagai kemungkinan penulisan suatu aksara. Setiap hasil pengolahan sampel akan disimpan kedalam sebuah basis data, sehingga dapat dibuka dan diperbaharui data tersebut.

Adapun data pengetahuan yang tersimpan pada basis data pada tahap pelatihan ini dapat dilihat pada Gambar 5. Dapat dilihat, terdapat *cell* berwarna kuning yang merupakan nilai 1 yang merupakan warna hitam pada piksel gambar yang sudah disesuaikan ukurannya dengan menggunakan tahap 2 pada tahap pelatihan. Basis data ini dapat terus diperbaharui dengan melatih sistem pengenalan aksara dengan karakter aksara baru yang memiliki cara penulisan yang berbeda sehingga sistem semakin kaya akan informasi berbagai macam aksara Kaganga Lampung.

Selanjutnya pada tahap pengujian, sistem pengenalan aksara kaganga Lampung diuji dengan 10 data penulisan yang berbeda dengan hasil pelatihan, adapun data uji dan hasil pengujian ditunjukkan pada Tabel 1.

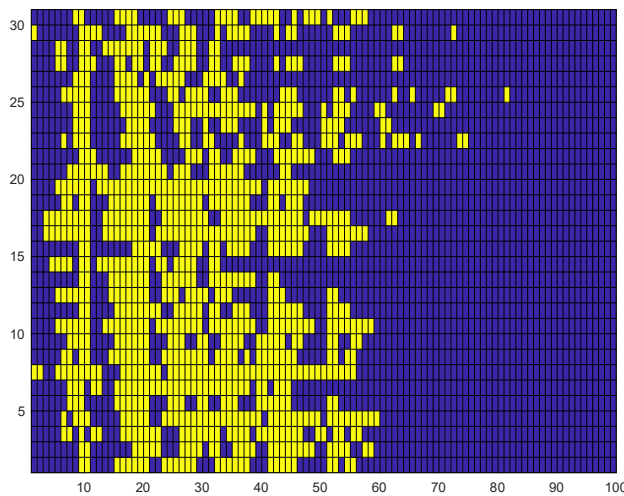
**Tabel 1.** Data Hasil Pengujian Pertama

No	Aksara	Gambar Uji	Waktu (s)	Sukses	Dikenali Sebagai
1	ka		1.1867	T	nga

2	ga		0.8931	Y	ga
3	nga		1.2164	T	nya
4	ba		1.0846	Y	ba
5	ma		0.9298	Y	ma
6	ta		1.0780	T	ka
7	ca		1.1325	T	ka
8	ja		0.9516	T	na
9	ya		0.6536	Y	ya
10	gha		1.2534	T	ma

Dari hasil beberapa pengujian, dengan basis data berjumlah 20 karakter, persentase keberhasilan hasil yang diperoleh pengenalan aksara sebesar 40% dengan rata-rata waktu proses pengenalan aksara sebesar 1.0380 detik. Nilai akurasi pengenalan aksara ini masih tergolong sangat rendah, hal ini dapat disebabkan oleh dataset yang digunakan pada pelatihan dan data uji memiliki perbedaan yang belum mampu dikenali oleh sistem pengenalan aksara ini. Untuk mengatasi hal tersebut perlu dilakukan pelatihan data tambahan dengan menggunakan gambar aksara data uji yang baru, hal ini bertujuan untuk memperkaya sistem dengan data pengetahuan baru.

Setelah sistem pengenalan aksara kaganga Lampung dilatih ulang dengan basis data baru seperti pada Gambar 6 dapat dilihat bahwa terdapat penambahan jumlah data hasil latihan yang menunjukkan data latihan baru sudah tersimpan kedalam basis data pengetahuan. Langkah selanjutnya adalah menguji aksara dengan gambar uji baru. Hasil pengujian dengan data baru ditunjukkan pada Tabel 2 berikut ini.



**Gambar 6.** Basis data pengetahuan setelah dilatih dengan data aksara baru

Dari pengujian kedua ini menunjukkan bahwa sistem pengenalan aksara Kaganga Lampung memiliki performa yang lebih baik dalam hal akurasi ketika dilatih ulang dengan menambahkan variasi bentuk aksara yaitu diperoleh nilai akurasi sebesar 70%. Namun dengan bertambahnya data pengetahuan pada basis data membuat sistem ini memiliki proses yang sedikit lebih lambat dibandingkan sebelumnya dengan waktu proses rata-rata sebesar 1.1412 detik. Hal ini dikarenakan dengan semakin banyak data pengetahuan, maka metode K-NN harus memeriksa data latih yang lebih banyak. Proses peningkatan waktu proses ini dapat semakin lebih lambat ketika basis data pengetahuan yang digunakan semakin besar.

**Tabel 2.** Hasil Pengujian Kedua

No	Aksara	Gambar Uji	Waktu (s)	Sukses	Dikenali Sebagai
1	ka		1.2027	Y	ka
2	ga		0.7845	Y	ga
3	nga		1.2635	Y	nga
4	ba		1.0195	T	pa
5	ma		1.0078	Y	ma
6	ta		0.9267	T	ga

7	ca		1.2112	Y	ca
8	ja		0.9411	Y	ja
9	ya		1.1985	T	pa
10	gha		1.8566	Y	gha

#### IV. KESIMPULAN

Sistem pengenalan aksara kaganga Lampung pada manuskrip ini menggunakan Metode K-Nearest neighbour (K-NN). Metode ini membutuhkan dua tahap yaitu tahap pelatihan dan tahap pengujian. Setelah sistem dilatih dan kemudian diuji, menunjukkan bahwa akurasi sistem masih rendah yaitu sekitar 40%, setelah dilakukan pelatihan dengan memberikan gambar 10 aksara yang baru, menunjukkan bahwa Sistem pengenalan aksara dengan metode K-NN memiliki peningkatan akurasi menjadi 70%, namun proses pengenalan menjadi sedikit lebih lambat karena jumlah data latih juga semakin banyak untuk diperiksa. Hal ini memberikan kesempatan atau celah penelitian selanjutnya terkait mempercepat proses pengenalan meskipun data latih semakin banyak.

#### REFERENSI

- [1] U. A. T. . Duludu, “Model Pelatihan Berbasis Kearifan Lokal Dalam Meningkatkan Profesionalisme Tutor Paket C Di Kabupaten Bone Bolango,” *J. Bakti Sar.*, vol. 5, no. 1, pp. 9–16, 2012.
- [2] G. Hildred; R. A. Zainuddin, T.O. Ihroni. “*Aneka budaya dan komunitas di Indonesia*” Jakarta :: Yayasan Ilmu-Ilmu Sosial, 1981.
- [3] A. Budiman, P. W. Wijaya, M. H. A. Nur, M. V. Halawa, and S. Susyanti, “Revisualisasi Aksara Kaganga Sebagai Identitas Lampung,” *J. Bhs. Rupa*, vol. 2, no. 2, pp. 118–125, 2019, doi: 10.31598/bahasarupa.v2i2.344.
- [4] R. I. Indrayati and N. Migotuwio, “Identifikasi Anatomi Aksara Lampung,” *AKSA J. Desain*

Komun. Vis., vol. 4, no. 1, pp. 541–551, 2020, doi: 10.37505/aksa.v4i1.43.

- [5] D. Syaputra, “Aksara Incung dalam Naskah di Kerinci (Incung Script and Manuscript In Kerinci),” *Hadharah J. Keislam. dan Perad.*, vol. 15, no. 1, pp. 71–90, 2021.
- [6] H. Mubarat, “JURNAL EKSPRESI SENI Jurnal Ilmu Pengetahuan dan Karya Seni,” *J. Ilmu Pengetah. dan Karya Seni*, vol. 17, no. 2, pp. 165–179, 2015.
- [7] R. M. Ridwan, A. Z. Mansoor, and I. S. Mustikadara, “Perancangan Game Puzzle sebagai Media Pembelajaran Aksara Sunda untuk Remaja,” *Wimba J. Komun. Vis.*, vol. 5, no. 1, pp. 27–42, 2015, doi: 10.5614/jkvw.2013.5.1.3.
- [8] R. Ptucha, F. P. Such, S. Pillai, F. Brockler, V. Singh and P. Hutkowski, "Intelligent character recognition using fully convolutional neural networks", *Pattern Recognition*, vol. 88, pp. 604-613, 2019.
- [9] R. Sharma, B. Kaushik and N. K. Gondhi, "Devanagari and Gurmukhi Script Recognition in the Context of Machine Learning Classifiers", *Journal of Artificial Intelligence*, vol. 11, no. 2, pp. 65-70, 2018.
- [10] P. Sahare and S.B. Dhok, "Multilingual character segmentation and recognition schemes for Indian document images", *IEEE Access*, vol. 6, pp. 10603-10617, 2018.
- [11] R. Pramanik and S. Bag, "Shape decomposition-based handwritten compound character recognition for bangla ocr", *Journal of Visual Communication and Image Representation*, vol. 50, pp. 123-134, 2018.
- [12] Ong. Veronica, and D. Suhartono. "Using K-Nearest Neighbor in Optical Character Recognition." *ComTech*, vol. 7, no. 1, pp. 53-65. 2016.