ARTIKEL

PENGENALAN TULISAN TANGAN AKSARA THAI MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)

Oleh:

TIYAS AYU CATUR PRIHATINY

14.1.03.02.0314

Dibimbing oleh :

1. **Intan Nur Farida, M.Kom**
2. **Ratih Kumalasari Niswatin, S.ST., M.Kom**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

TAHUN 2018

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2018

Yang bertanda tangan di bawah ini:

Nama Lengkap	: Tiyas Ayu Catur Prihatiny
NPM	: 14.1.03.02.0314
Telepun/HP	: 0857 3679 0120
Alamat Surel (Email)	: ayucatur1495@gmail.com
Judul Artikel	: Pengenalan Tulisan Tangan Aksara Thai Menggunakan Metode *K-Nearest Neighbor* (K-NN)
Fakultas – Program Studi	: Teknik – Teknik Informatika
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi	: Jl. K.H. Achmad Dahlan No.76 Telepon (0354) 771503 Kota Kediri

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 7 Agustus 2018
Pembimbing I	Pembimbing II	Penulis,
Intan Nur Farida, M.Kom	**Ratih Kumalasari Niswatin, S.ST., M.Kom**	**Tiyas Ayu Catur P.**
NIDN. 0704108701	NIDN. 0710018501	NPM. 14.1.03.02.0314

PENGENALAN TULISAN TANGAN AKSARA THAI MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)

Tiyas Ayu Catur Prihatiny
14.1.03.02.0314
Fakultas Teknik – Teknik Informatika
ayucatur1495@gmail.com
Intan Nur Farida, M.Kom dan Ratih Kumalasari Niswatin, S.ST., M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Tulisan adalah bentuk secara tertulis dari bahasa. Salah satunya adalah aksara Thai yang digunakan untuk menulis Bahasa Thailand. Aksara Thai merupakan salah satu aksara yang sulit untuk dipelajari. Salah satu cara untuk mempelajarinya adalah dengan mengenal dasar aksara yang digunakan dan kemudian mencontohnya sebagai praktik menulis. Namun, tulisan tangan seseorang tidak selalu sama persis sekalipun telah diberikan pola huruf yang paten. Hal inilah yang nantinya akan memberikan hasil penilaian yang berbeda.

Permasalahan penelitian ini adalah tentang cara merancang dan membangun sistem yang dapat digunakan sebagai pengenalan aksara Thai, dan menerapkan metode *K-Nearest Neighbor* (K-NN) untuk melakukan klasifikasi pada aksara Thai. Dan penggunaannya menggunakan metode Integral Proyeksi sebagai pengekstraksi fitur, dan menggunakan *Euclidean Distance* untuk menghitung jarak. Sebelum dilakukan pengenalan, *file* citra dilakukan *preprocessing* terlebih dahulu, yaitu *grayscalling* dan biner, untuk selanjutnya didapatakn fitur, lalu diidentifikasi.

Dari hasil pengujian pada skenario pertama dengan data *training* 48 dan *testing* 72, diperoleh akurasi sebesar 42,2%. Pada skenario kedua dengan data *training* 72 dan *testing* 48, diperoleh akurasi sebesar 68,8%. Pada skenario ketiga dengan data *training* 120 dan *testing* 72, diperoleh akurasi sebesar 80,5%. Dari hasil pengujian, dapat disimpulkan bahwa metode *K-Nearest Neighbor* dan Integral Proyeksi dapat digunakan untuk mengidentifikasi tulisan tangan huruf Thai.

Kata kunci : Aksara Thai, Integral Proyeksi, *K-Nearest Neighbor* (K-NN)

I. LATAR BELAKANG

Dalam menjalin hubungan erat antar setiap manusia dibutuhkan media penghubung untuk dapat saling mengutarakan pendapat dan pikiran. Di mana terdapat metode komunikasi secara lisan dan tertulis. Dalam hal ini mengambil media komunikasi secara tertulis. “Tulisan merupakan sebuah alat untuk mengungkapkan ide, gagasan dan pikiran kita kepada orang lain secara tertulis” (Tarigan, 1986).

“Aksara Thai merupakan sebagian huruf yang diambil dari China, beberapa dari Melayu, dan beberapa ratus lebih dari Pali, Sansekerta, dan Khmer” (Jacob Y. de Groot, 2010). Aksara Thai termasuk dalam salah satu aksara yang sulit dipelajari. Apalagi bagi para pemula yang baru akan belajar menulis

aksara Thai. Seperti bagi mereka yang memutuskan untuk melanjutkan pendidikan maupun bekerja ke Negeri Gajah Putih tersebut pasti membutuhkan sarana pembelajaran, di mana kursus Bahasa Thai sendiri masih sangat jarang di Indonesia. Pada pengolahan citra terdapat pengenalan pola di mana hal ini dapat digunakan untuk mengenali huruf tulisan dari aksara Thai sebagai media pembantu pembelajaran.

Penelitian tentang pengenalan pola pada huruf telah banyak dilakukan sebelumnya, salah satunya ialah Pengenalan Huruf Arab Menggunakan Metode Reduksi *Two Dimensional Principal Component Analysis* (2DPCA) dan Klasifikasi *K-Nearest Neighbor* (K-NN) oleh Masitoh Majid, Arid F. Huda, Rini Cahyandari (2016) dengan tingkat keberhasilan tertinggi mencapai 95% dengan menggunakan metode *Bilateral 2 Dimensional Principal Component Analysis* (B2DPCA).

Dari beberapa uraian di atas menjadikan penulis tertarik untuk membuat suatu aplikasi "Pengenalan Tulisan Tangan Aksara Thai Menggunakan Metode *K-Nearest Neighbor* (K-NN)" untuk pengenalan tulisan tangan aksara Thai yang sesuai dan optimal.

II. METODE

A. *K-Nearest Neighbor* (K-NN)

"*K-Nearest Neighbor* (K-NN) adalah suatu metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan data *training* yang jaraknya paling dekat dengan obyek tersebut" (Mitchell, 1997). "Prinsip kerja K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *k* tetangga (*neighbor*) terdekatnya dalam pelatihan" (Helilintar, 2017: 34).

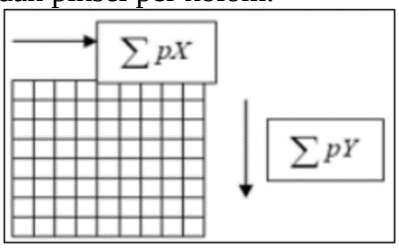
Menurut Gorunescu (2011), langkah-langkah dalam implementasi perhitungan metode K-NN adalah sebagai berikut :

1. Menentukan parameter *k* (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclidean* (*Euclidean Distance*) masing-masing obyek terhadap data sampel yang diberikan.
3. Mengurutkan jarak yang terbentuk (urut naik).
4. Memilih *alternative* jarak terdekat sebanyak *k*.

B. Integral Proyeksi

Menurut Subiyantoro dan Putra (2011), integral proyeksi adalah suatu metode yang digunakan untuk mencari daerah atau lokasi dari obyek. Metode ini dapat digunakan untuk mendeteksi batas dari daerah

gambar yang berbeda, sehingga kita bisa mencari daerah lokasi, huruf, dan fitur-fiturnya. Metode ini juga bisa disebut dengan integral baris dan kolom dari piksel, karena integral ini menjumlahkan piksel per baris dan piksel per kolom.



Gambar 1 : Integral Proyeksi

Integral proyeksi secara umum didefinisikan seperti ditunjukkan pada Persamaan 1 dan Persamaan 2 (Amaliah, Yuniarti, Nugroho, & Arifin, 2011) :

$h(i) = \sum_{i=1}^{N\ baris} x(i,j)$..... (1)

$h(j) = \sum_{j=1}^{N\ kolom} x(i,j)$..... (2)

Di mana *h(i)* merupakan penghitungan integral secara vertikal, sedangkan *h(j)* merupakan penghitungan integral secara horisontal.

C. *Euclidean Distance*

"Jarak menjadi aspek penting dalam pengembangan metode pengklasifikasian maupun regresi" (Santoso, 2007).

Menurut Darma Putra (2010), *euclidean distance* dimaknai sebagai berikut :

> *Euclidean Distance* adalah matriks yang paling sering diguankan untuk menghitung kesamaan 2 vektor. Jarak *euclidean* menghitung akar dari kuadarat perbedaan 2 vektor (*root of square differences between 2 vector*). Persamaan dari jarak *euclidean* adalah :

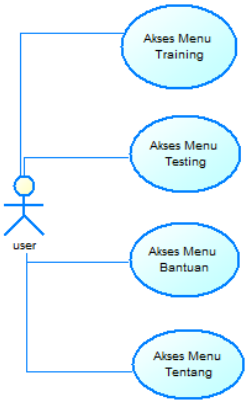
$$d_{ij} = \sqrt{\sum_{k=1}^{n}\left(x_{ik} - x_{jk}\right)^2} \text{.....(2.3)}$$

Hasil perhitungan *euclidean distance* ini akan memperlihatkan seberapa besar tingkat kesamaan antara citra uji dan citra sampel, semakin kecil nilai dari *euclidean distance* (mendekati nilai nol), maka tingkat kemiripan (*similarity*) citra semakin baik

III. HASIL DAN KESIMPULAN

A. Implementasi Sistem

Use Case Diagram



Gambar 2 : *Use Case Diagram*

Keterangan *use case diagram* pada gambar 2 adalah sebagai berikut :

1) Akses menu *training* yang di dalamnya melibatkan proses sebagai berikut :
   a) Pengambilan citra *training*. Citra *training* yang diambil

adalah huruf Thai **Kh** (ค), **Kh** (ฅ), **Kh** (ฆ), **Ng** (ง), **Ch** (ช), **S** (ซ), **Ch** (ฌ), **Y** (ญ), **Th** (ฑ), **Th** (ฒ), **N** (ณ), **Th** (ท), **Th** (ธ), **N** (น), **Ph** (ผ), **F** (ฝ), **Ph** (ภ), **M** (ม), **Y** (ย), **R** (ร), **L** (ล), **W** (ว), **L** (ฬ), **H** (ฮ). Aksara Thai tersebut diambil dari "PT. Ruang Bahasa Internasional".

b) *Preprocessing* ini melakukan pengambilan nilai ekstraksi fitur dengan menggunakan integral proyeksi.

c) *Database* digunakan untuk menyimpan citra *training* yang dapat dijadikan acuan untuk melakukan identifikasi atau pengenalan tulisan tangan aksara Thai.

2) Akses menu *testing* yang di dalamnya melibatkan proses sebagai berikut :

a) Pengambilan citra *testing*. Citra *testing* yang diambil adalah huruf Thai **Kh** (ค), **Kh** (ฅ), **Kh** (ฆ), **Ng** (ง), **Ch** (ช), **S** (ซ), **Ch** (ฌ), **Y** (ญ), **Th** (ฑ), **Th** (ฒ), **N** (ณ), **Th** (ท), **Th** (ธ), **N** (น), **Ph** (ผ), **F** (ฝ), **Ph** (ภ), **M** (ม), **Y** (ย), **R** (ร), **L** (ล), **W** (ว), **L** (ฬ), **H** (ฮ). Aksara Thai tersebut diambil dari "PT. Ruang Bahasa Internasional".

b) *Preprocessing* ini melakukan pengambilan nilai ekstraksi fitur dengan menggunakan integral proyeksi.

c) Proses pencocokan menggunakan metode *euclidean distance*. Metode ini digunakan untuk melakukan perhitungan kedekatan jarak antara citra *training* dengan citra *testing*.

d) Hasil identifikasi citra huruf merupakan hasil identifikasi yang akan ditampilkan berupa nama huruf yang berhasil dikenali.

3) Akses menu bantuan untuk mendapatkan penjelasan tentang setiap *form* yang ada di aplikasi.

4) Akses menu tentang untuk mendapatkan informasi tentang aplikasi.

B. Tampilan Program

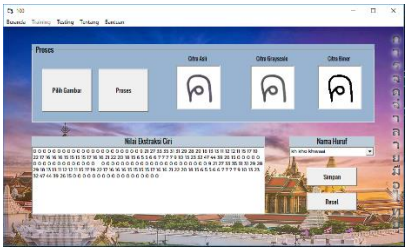
a. Tampilan Beranda

Beranda adalah *form* yang pertama kali muncul saat program dijalankan. Tampilan beranda dapat dilihat pada gambar 3.

Gambar 3 : Tampilan *Form* Beranda

b. Tampilan *Training*

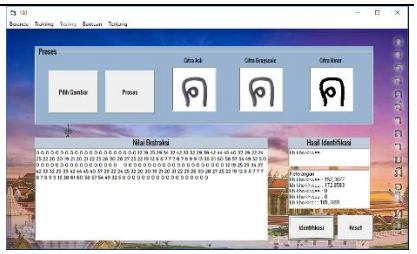
Training adalah *form* yang dapat diakses dari *form* beranda. *Form* ini digunakan sebagai akses untuk memproses data *training*. Tampilan *training* dapat dilihat pada gambar 4.



Gambar 4 : Tampilan *Form Training*

c. Tampilan *Testing*

Testing adalah *form* yang dapat diakses dari *form* beranda. *Form* ini digunakan sebagai proses identifikasi huruf. Tampilan *testing* dapat dilihat pada gambar 5.



Gambar 5 : Tampilan *Form Testing*

C. Skenario Uji Coba

Dari hasil uji coba 120 data *testing* didapatkan hasil yang terlihat pada tabel 1 sebagai berikut:

No	Data *Training*	Data *Testing*	Benar	Salah	Akurasi
1	48 citra sampel dari 24 aksara Thai.	72 citra sampel dari 24 aksara Thai.	34	38	$34 \times \frac{100}{72} = 42{,}2\%$
2	72 citra sampel dari 24 aksara Thai.	48 citra sampel dari 24 aksara Thai.	33	15	$33 \times \frac{100}{48} = 68{,}8\%$
3	120 citra sampel dari 24 aksara Thai.	72 citra sampel dari 24 aksara Thai.	58	14	$58 \times \frac{100}{72} = 80{,}5\%$

Dari data di atas dapat disimpulkan bahwa hasil tertinggi sebesar 80,5%. %. Di mana pada uji coba ini menggunakan citra training yang lebih banyak daripada dua skenario lainnya. Dengan kata lain semakin banyak data training yang digunakan akan semakin menambah nilai keberhasilan yang akan didapatkan dalam mendeteksi aksara Thai.

IV. PENUTUP

A. Simpulan

Dari hasil penelitian, perancangan, pembuatan, dan pengujian aplikasi pengenalan tulisan tangan aksara Thai menggunakan metode *K-Nearest Neighbor* (K-NN) didapatkan kesimpulan sebagai berikut :

1. Metode *K-Nearest Neighbor* (K-NN) dapat diguanakn untuk mengidentifikasi aksara Thai. Di mana pada aplikasi ini bentuk memiliki peran penting dalam proses identifikasi aksara Thai. Semakin rapi bentuk tulisan tangan pada lembar pengambilan sampel dengan memperhatikan garis bantu dan garis tepi yang tersedia, dengan kata lain semakin mirip tulisan tangan dengan tulisan asli yang telah dipatenkan sebagai contoh maka presentase tingkat keberhasilan yang akan didapatkan semakin tinggi.
2. Tingkat akurasi identifikasi dengan benar menggunakan tahapan penelitian tersebut dan menggunakan 120 data uji pada citra yang diambil adalah 80,5%.

B. Saran

Dari hasil penelitian, perancangan, pembuatan, dan pengujian aplikasi pengenalan tulisan tangan aksara Thai menggunakan metode K-Nearest Neighbor (K-NN) didapatkan saran untuk penelitian selanjutnya sebagai berikut :

1. Penelitian dilakukan dengan kasus yang berbeda dari pengenalan tulisan tangan aksara Thai, dengan menggunakan huruf konsonan kelas tinggi maupun kelas tengah.
2. Menambah jumlah percobaan di dalam pengenalan tulisan tangan aksara Thai dengan jumlah data vektor yang berbeda agar keakuratan nilai yang lebih baik.
3. Penelitian selanjutnya dapat menggunakan gabungan metode yang lain guna memperoleh hasil yang lebih baik bila dibandingkan dengan 2 metode yang digunakan pada penelitian ini.
4. Lebih memperhatikan citra yang akan digunakan. Kesesuaian ukuran, jarak tepi ke tulisan juga mempengaruhi hasil.

V. DAFTAR PUSTAKA

Amaliah, B., Anny Yuniarti, Anindita Sigit Nugroho, dan Agus Zainal Arifin. 2011. *Pemisahan Gigi pada Dental Panoramic Radiograph dengan Menggunakan Integral Projection yang Dimodifikasi*. (Online) Vol. 6, No. 2, 121-128, tersedia: http://journal.trunojoyo.ac.id/kursor/article/view/1116, diunduh 5 Juli 2018.

Gorunescu, Florin. 2011. *Data Mining: Concepts, Models and Techniques*. Springer.

Groot, Jacob Y. de. 2010. *Thai for Beginners*. Thailand: Prince of Songkhla University.

Helilintar, Risa, Risky Aswi Ramadhani, dan Siti Rochana. 2017. *Data Mining: K-Nearest Neighbor* (Danar Putra Pamungkas, Ed.). Kediri: Fakultas Teknik Informatika Nusantara PGRI.

Majid, Masitoh, Arif F. Huda dan Rini Cahyandari. 2016. *Pengenalan Huruf Arab Menggunakan Metode Reduksi Two Dimensional Principal Component Analysis (2DPCA) dan Klasifikasi K-Nearest Neighbor (K-NN)*. (Online) 27 – 34, tersedia: http://jurnal.unpad.ac.id/jmi/article/view/10281, diunduh 22 Agustus 2017.

Mitchell, T. 1997. *Machine Learning*. McGraw-Hill, New York.

Putra, Darma. 2010. *Pengolahan Citra Digital*. Yogyakarta: Andi.

Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

Subiyantoro, E., dan Agung Prajuhana, P. 2016. *Penentuan Kualitas Daun Tembakau dengan Perangkat Mobile Berdasarkan Ekstraksi Fitur Rata-rata RGB Menggunakan Algoritma K-Nearest Neighbor*. (Online), tersedia: https://www.researchgate.net/publication/267722708, diunduh 5 Juli 2018.

Tarigan, Djago. 1986. *Teknik Pengajaran Keterampilan Berbahasa*. Bandung: Angkasa.