

Perbandingan Akurasi Algoritma K-Nearest Neighbors (KNN) dan Naive Pattern Search (NPS) dalam Website Handwritten Recognition untuk Latihan Menulis Bentuk

Ni Putu Linda Santiari¹, I Gede Surya Rahayuda²

¹Fakultas Informatika dan Komputer, Program Studi Sistem Informasi
Institut Teknologi dan Bisnis STIKOM Bali
Denpasar, Indonesia

²Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Studi Informatika
Universitas Udayana
Mangupura, Indonesia

e-mail: ¹linda_santiari@stikom-bali.ac.id, ²igedesuryarahayuda@unud.ac.id

Abstrak

Pengenalan tulisan tangan (handwritten recognition) menjadi semakin relevan dalam era perkembangan teknologi saat ini. Dalam artikel ini, membahas tentang perbandingan akurasi algoritma KNN dan NPS yang dipergunakan dalam pengembangan website handwritten recognition khusus untuk latihan menulis bentuk bagi anak-anak usia 4-7 tahun. perbandingan antara dua algoritma yang digunakan dalam pengenalan angka tulisan tangan, yaitu algoritma K-Nearest Neighbors (KNN) dan Naive Pattern Search (NPS). KNN adalah algoritma klasifikasi yang populer, yang menggunakan mayoritas dari kategori tetangga terdekat untuk mengklasifikasikan data uji. Sementara itu, NPS adalah algoritma pencocokan pola sederhana yang mencari kesamaan dengan contoh-contoh referensi dalam data sampel. Data pengujian diambil dari hasil gambar png yang disimpan saat uji coba website. Kriteria pengujian adalah akurasi hasil yang diperoleh dari data test. Hasil gambar data test kemudian disimpan dan diproses recognisi dengan algoritma dan melihat persentase hasil kemiripan dengan data sample pada website. Hasil dari perbandingan kedua algoritma menunjukkan bahwa keduanya memberikan hasil yang baik dalam pengenalan bentuk tulisan tangan. Akurasi yang dicapai oleh algoritma KNN mencapai 70%, sementara NPS mencapai 68%. Akurasi diperoleh dengan perhitungan manual. Meskipun perbedaan akurasi tidak signifikan, hal ini dapat dipengaruhi oleh variasi gaya tulisan tangan dari masing-masing siswa yang mencoba latihan pada website.

Kata kunci: handwritten, recognition, KNN, NPS

Abstract

Handwritten recognition is becoming increasingly relevant in the current era of technological development. In this article, we discuss the comparison of the accuracy of the KNN and NPS algorithms used in developing handwritten recognition websites specifically for practicing writing shapes for children aged 4–7 years. comparison between the two algorithms used in the recognition of handwritten numbers, namely the K-Nearest Neighbors (KNN) and Naive Pattern Search (NPS) algorithms. KNN is a popular classification algorithm that uses the majority of nearest neighbor categories to classify test data. Meanwhile, NPS is a simple pattern matching algorithm that looks for similarities to reference examples in sample data. The test data is taken from the results of the PNG images stored during the website trial. The test criterion is the accuracy of the results obtained from the test data. The results of the test data images are then stored and processed for recognition with algorithms, and you can see the percentage of similarity results with the sample data on the website. The results of the comparison of the two algorithms show that both of them provide good results in handwriting recognition. The accuracy achieved by the KNN algorithm reaches 70%, while the NPS reaches 68%. Accuracy is obtained by manual calculation. Although the difference in accuracy is not significant, it can be influenced by variations in the handwriting style of each student who tries exercises on the website.

Keywords: handwritten, recognition, KNN, NPS

1. Pendahuluan

Perkembangan teknologi pada era ini telah mengalami lonjakan yang luar biasa, membawa dampak besar pada hampir setiap aspek kehidupan manusia. Kemajuan teknologi informasi dan komunikasi telah mengubah cara kita berinteraksi, belajar, bekerja, dan berkomunikasi. Revolusi digital telah membuka pintu untuk inovasi baru, menghadirkan berbagai aplikasi yang mendukung aktivitas sehari-hari dengan tingkat efisiensi yang lebih tinggi.

Salah satu bidang yang telah mengalami transformasi signifikan adalah pengenalan tulisan tangan (*handwritten recognition*)[1]. Dalam penelitian ini yaitu *website handwritten recognition* untuk Latihan menulis bentuk. Website ini adalah website yang dapat dipergunakan oleh anak-anak dalam Latihan menulis bentuk bidang sederhana. Dalam website ini Latihan menulis yang dapat dilakukan yaitu bentuk lingkaran, persegi, segitiga, belah ketupat, dan bintang. Tujuan dari dibangunnya website ini yaitu untuk meningkatkan Latihan menulis tangan pada anak usia 4-7 tahun dengan menggunakan media digital. *Handwritten recognition* dikembangkan berbasis web menggunakan *framework bootstrap* dan menggunakan algoritma KNN dan *Naive Pattern Search*.

Latihan menulis bentuk dengan *website handwritten recognition* untuk Latihan menulis bentuk ini perlu adanya perbandingan akurasi algoritma yang dipergunakan. Dari permasalahan tersebut perlu adanya perbandingan untuk algoritma untuk website ini. Beberapa penelitian sebelumnya tentang *handwritten* yaitu Penelitian sebelumnya dengan judul Implementasi *Zoning* dan Fitur Arah Sebagai Ekstraksi Fitur, Dalam studi ini dibahas mengenai penggabungan pendekatan yang bergantung pada zonasi dan fitur arah. Metode ICZ, ZCZ, dan *freeman chain code* digunakan untuk menentukan ciri-ciri tulisan tangan aksara Bali. Penambahan pendekatan kode rantai mencoba menemukan makna di sekitar titik latar depan. Efek ekstraksi fitur akan digunakan sebagai umpan balik dalam proses klasifikasi oleh *Support Vector Machine*. Hasil penelitian menunjukkan bahwa menggabungkan metode ICZ, ZCZ, dan *freeman chain code* menghasilkan akurasi 89,09 persen, sedangkan kombinasi ICZ dan ZCZ menghasilkan akurasi 88,06 persen. *Kernel linier* termasuk dalam kernel SVM yang dikontraskan [2].

Penelitian kedua dengan judul *Arabic Handwritten Characters Recognition using Convolutional Neural Network*. Dalam penelitian ini, penulis mengembangkan arsitektur deep learning yang dapat efektif digunakan untuk mengenali karakter Arab tulisan tangan. Mereka menggunakan *Convolutional Neural Network (CNN)*, yang merupakan jenis khusus dari jaringan saraf tiruan berlapis maju yang dilatih dalam mode terpantau. Pada penelitian ini dibahas penggunaan algoritma CNN dalam pengenalan karakter tulisan tangan Arab. Penggunaan CNN menghasilkan peningkatan yang signifikan dalam berbagai algoritma klasifikasi pembelajaran mesin. CNN yang diusulkan memberikan rata-rata 5,1% kesalahan klasifikasi pada data pengujian[3].

Penelitian ketiga dengan judul *CNN-based multilingual handwritten numeral recognition: A fusion-free approach*. Dalam penelitian ini, penulis telah mengembangkan sistem pengenalan angka tulisan tangan yang independen dari aksara untuk angka multibahasa, yang tidak memerlukan penggabungan kelas dan hanya memiliki 10 kelas untuk setiap angka numerik tunggal. Ini merupakan karya pertama dalam menangani masalah sistem pengenalan angka multibahasa. Diperoleh akurasi 96,23% secara kolektif untuk kedelapan aksara tersebut. Akurasi tinggi ini menunjukkan bahwa pengenalan angka tulisan tangan multibahasa dapat dilakukan dengan sukses menggunakan *Convolutional Neural Network (CNN)*[4]. Dari beberapa penelitian tersebut, peneliti mempunyai solusi terhadap permasalahan yang ditemukan di lapangan yaitu membandingkan akurasi metode KNN dan NPS pada *website handwritten recognition* untuk latihan menulis bentuk. Hasil dari perbandingan ini diharapkan dapat lebih meningkatkan akurasi hasil dari website ini.

2. Metode Penelitian

Untuk klasifikasi matrik gambar digunakan algoritma KNN dan *Naive Pattern Search*. Data yang digunakan pada pengujian ini yaitu data gambar yang tersimpan saat uji coba website. Sepuluh buah gambar sampel yang ada kemudian dirubah menjadi matrik dua dimensi, dimana terdapat perbedaan nilai antara gambar kosong dan gambar yang terisi goresan layar sentuh. Misalkan pixel dari gambar tanpa goresan diberi nilai 0, sedangkan pixel dari gambar yang terisi goresan diberi nilai dari 1-3 berdasarkan ketebalan goresan atau tekanan pada saat menulis pada layar. Begitu juga dengan data tes yang baru dituliskan, gambar tersebut juga dikonversi menjadi matrik dua dimensi. Kemudian data tes tersebut dibandingkan menggunakan algoritma KNN dan NPS.

K-Nearest Neighbors (k-NN) adalah salah satu algoritma klasifikasi yang populer dalam pengenalan pola dan analisis data. Algoritma ini digunakan untuk mengklasifikasikan objek atau data berdasarkan mayoritas dari kategori tetangga terdekatnya. Prinsip dasar dari k-NN adalah bahwa objek dengan atribut yang mirip cenderung berada dalam kelas atau kategori yang sama[5].

Proses k-NN dimulai dengan memuatkan data sampel yang berisi contoh-contoh dengan atribut dan labelnya ke dalam struktur data yang memungkinkan pencarian tetangga terdekat dengan cepat, seperti KD-Tree atau Ball Tree[6]. Ketika ada data uji yang akan diklasifikasikan, algoritma k-NN akan mencari k tetangga terdekat dari data uji berdasarkan fungsi jarak tertentu, misalnya jarak Euclidean atau jarak Manhattan. Setelah k tetangga terdekat ditemukan, data uji akan diberikan label atau kelas berdasarkan mayoritas dari label tetangga terdekatnya. Jika nilai $k = 1$, maka data uji akan diberikan label dari satu tetangga terdekatnya[7].

Kelebihan dari algoritma k-NN adalah kemampuannya untuk mengatasi masalah klasifikasi non-linear dan mampu mengenali pola yang kompleks[8]. Namun, algoritma ini memiliki kelemahan dalam hal efisiensi komputasi, terutama ketika jumlah data pelatihan sangat besar. Selain itu, nilai k yang digunakan dalam k-NN harus dipilih secara bijaksana, karena nilai k yang terlalu kecil dapat menyebabkan model menjadi sensitif terhadap noise, sementara nilai yang terlalu besar dapat menyebabkan model menjadi terlalu umum dan kurang responsif terhadap variasi data[9].

Naive Pattern Search adalah algoritma pencocokan pola sederhana yang digunakan untuk mengenali pola dari data uji dengan mencari kesamaan dengan contoh-contoh referensi dalam database. Algoritma ini merupakan pendekatan yang naif karena tidak melibatkan proses pembelajaran dari data sampel[10]. Dalam Naive Pattern Search, setiap karakter atau tulisan tangan akan direpresentasikan sebagai vektor fitur yang menggambarkan atribut-atribut tertentu dari karakter tersebut. Misalnya, atribut-atribut tersebut dapat berupa bentuk garis, sudut-sudut karakter, atau panjang lekukan. Data uji akan dibandingkan secara langsung dengan contoh-contoh referensi dalam database, kemudian mencari kesamaan pola atau kemiripan yang cukup tinggi. Jika ditemukan kesamaan pola yang signifikan, maka data uji akan diklasifikasikan dengan label yang sesuai dengan contoh referensi yang paling mirip. Pada penelitian ini dilakukan perbandingan untuk membandingkan hasil yang telah didapatkan pada proses implementasi antara algoritma KNN dan *Naive Pattern Search*.

3. Hasil dan Pembahasan

3.1 Data

Pengambilan data dilakukan dengan memberikan link berupa text url atau QR code dari sistem yang telah dihosting pada web hosting kepada pihak sekolah. Link tersebut nantinya akan diberikan kepada Guru yang selanjutnya akan meneruskan kepada beberapa orangtua siswa untuk dicobakan kepada siswa. Siswa dapat mengakses sistem melalui link yang diberikan dan memulai uji coba latihan menulis. Media yang digunakan dapat berupa smartphone atau tablet. Uji dilakukan pada 10 orang siswa, dimana tiap siswa melakukan uji coba pada 5 bentuk. Sehingga akan didapatkan sejumlah 5 gambar bentuk untuk setiap siswa. Total keseluruhan data menjadi 50 gambar angka untuk 10 siswa. Data sampel ini merupakan data gambar bentuk yang benar, data ini dapat diambil dari penulis sendiri, Guru atau mahasiswa. Pada tiap bentuk terdapat 10 data gambar, sehingga total terdapat sejumlah 50 data gambar untuk 5 bentuk. Teknik pengumpulan data yang digunakan adalah pengumpulan data studi dokumen, karena mengandalkan dokumen sebagai sumber data yang digunakan untuk penelitian. Dokumen yang digunakan berupa gambar lingkaran, persegi, segitiga, belahketupat dan bintang. Pengumpulan data dilakukan dengan memberikan link website kepada beberapa responden, dan kemudian ketika website diuji coba oleh responden, gambar lingkaran, persegi, segitiga, belahketupat dan bintang akan tersimpan pada web server.

3.2 Hasil Perbandingan Metode KNN dan NPS

Pengambilan data dilakukan dengan memberikan link berupa text url atau QR code kepada orangtua siswa. Link tersebut nantinya akan diberikan kepada Guru yang selanjutnya akan meneruskan kepada beberapa siswa. Siswa dapat mengakses sistem melalui link yang diberikan dan memulai uji coba latihan menulis shape. Media yang digunakan dapat berupa smartphone atau tablet. Uji coba akan dilakukan pada 10 orang siswa, dimana tiap siswa melakukan uji coba pada 5 bentuk bidang sederhana, dan tiap bentuk dilakukan uji coba 1 kali. Sehingga akan didapatkan sejumlah 5 gambar bentuk untuk setiap siswa. Total keseluruhan data menjadi 50 gambar huruf untuk 10 siswa. Rata umur siswa berkisar antara 4 – 7 tahun, dimana siswa tersebut sedang mendapatkan pelajaran menulis bentuk bidang sederhana.

Tabel 1. Data Tes dan Hasil

No	Nama	Bentuk				
		Lingkaran	Persegi	Segitiga	Belah Ketupat	Bintang
1	Agung					
	KNN Algorithm	40%	40%	40%	60%	100%
	NPS Algorithm	40%	20%	20%	40%	60%
2	Alana					
	KNN Algorithm	100%	40%	60%	80%	100%
	NPS Algorithm	80%	40%	40%	60%	60%
3	Atalia					
	KNN Algorithm	60%	80%	0%	80%	100%
	NPS Algorithm	60%	80%	0%	60%	60%
4	Ayu					
	KNN Algorithm	80%	20%	80%	100%	100%
	NPS Algorithm	20%	60%	80%	80%	80%
5	Geni					
	KNN Algorithm	40%	80%	60%	80%	100%
	NPS Algorithm	80%	80%	80%	20%	100%
6	Gungwik					
	KNN Algorithm	80%	80%	80%	60%	0%
	NPS Algorithm	80%	20%	60%	80%	0%

No	Nama	Bentuk				
		Lingkaran	Persegi	Segitiga	Belah Ketupat	Bintang
7	Nadia					
	KNN Algorithm	100%	60%	60%	100%	100%
	NPS Algorithm	80%	60%	60%	80%	80%
8	Risa					
	KNN Algorithm	0%	80%	0%	60%	100%
	NPS Algorithm	0%	60%	0%	40%	60%
9	Rizky					
	KNN Algorithm	20%	80%	20%	60%	100%
	NPS Algorithm	80%	60%	40%	60%	40%
10	Sanjaya					
	KNN Algorithm	40%	20%	100%	40%	100%
	NPS Algorithm	60%	60%	100%	80%	80%

Dari tabel tersebut diketahui bahwa semua siswa menuliskan angka dengan benar. Algoritma KNN terlihat memberikan nilai similarity yang cukup tinggi pada tiap gambar angka. Pada penelitian ini digunakan nilai $K=3$, diartikan bahwa data tes akan dikatakan benar ketika mendapatkan nilai sesuai minimal 6 dari 10 sampel atau bisa dikatakan nilai similarity minimal adalah 60%. Pada tabel 1 data tes dan hasil, dari 50 data nilai, terdapat sebanyak 35 gambar dengan nilai mencapai 60% atau lebih dan 15 gambar dengan nilai kurang dari 60%. Sehingga didapatkan nilai akurasi dari algoritma KNN sebesar 70%. Sedangkan untuk algoritma naïve pattern search, dari 50 data nilai, terdapat sebanyak 34 gambar dengan nilai mencapai 60% dan 16 gambar dengan nilai kurang dari 60%. Sehingga didapatkan nilai akurasi dari algoritma NPS sebesar 68%. Dapat disimpulkan bahwa algoritma KNN dan naïve pattern search berhasil dalam melakukan rekognisi. Jika diperhatikan pada tabel 5.5, dapat diketahui semua siswa menuliskan huruf dengan benar, namun beberapa tulisan mendapatkan hasil kurang dari 60%. Hal ini disebabkan karena pola tulisan tangan dari siswa berbeda dengan sepuluh contoh pada tabel 5.4. Beberapa siswa ada yang menuliskan bentuk kurang pas ditengah canvas, tulisan terlalu besar dan juga terlalu kecil. Hal tersebut tentunya akan mempengaruhi hasil rekognisi, karena akan menghasilkan matrik yang jauh berbeda dengan sampel. Algoritma naïve pattern search mendapatkan hasil yang kurang baik dibandingkan dengan

algoritma KNN, hal ini mungkin disebabkan karena algoritma naïve pattern dalam membandingkan matrik harus mencari nilai matrik yang benar – benar sama. Sehingga memiliki kemungkinan memberikan nilai tidak benar lebih banyak dibandingkan dengan algoritma KNN. Namun pada kasus penelitian ini algoritma KNN dan naïve pattern search hanya memiliki sedikit perbedaan nilai. Hal ini mungkin saja akan berbeda jika dilakukan pengujian pada jumlah data uji yang lebih banyak.

4. Kesimpulan

Website handwritten recognition untuk latihan menulis bentuk dibangun dengan algoritma KNN dan NPS. Akurasi terhadap kedua algoritma ini sudah dilakukan pada penelitian ini. Berdasarkan uji coba yang dilakukan oleh 10 siswa, didapatkan nilai akurasi dari algoritma KNN adalah 70% dan algoritma naïve pattern search adalah 68%. Nilai tersebut merupakan nilai yang cukup baik. Pengujian akurasi menggunakan perhitungan manual. Dapat disimpulkan algoritma KNN dan naïve pattern search berhasil melakukan recognisi terhadap gambar angka yang disesuaikan dengan data sampel Hal tersebut tentunya akan mempengaruhi hasil rekognisi, karena akan menghasilkan matrik yang jauh berbeda dengan sampel. Algoritma naïve pattern search mendapatkan hasil yang kurang baik dibandingkan dengan algoritma KNN, hal ini mungkin disebabkan karena algoritma naïve pattern dalam membandingkan matrik harus mencari nilai matrik yang benar – benar sama. Sehingga memiliki kemungkinan memberikan nilai tidak benar lebih banyak dibandingkan dengan algoritma KNN. Namun pada kasus penelitian ini algoritma KNN dan naïve pattern search hanya memiliki sedikit perbedaan nilai. Hal ini mungkin saja akan berbeda jika dilakukan pengujian pada jumlah data uji yang lebih banyak.

Daftar Pustaka

- [1] J. Brilliantio, N. Santosa, G. Ardian, and L. Hakim, "Penerapan convolutional neural network untuk handwriting recognition pada aplikasi belajar aritmatika dasar berbasis web," *Jurnal Teknik Informatika UNIKA Santo Thomas*, vol. 5, no. 2, pp. 137–146, 2020.
- [2] I. K. A. G. Wiguna and I. M. D. P. Asana, "Implementasi Zoning dan Fitur Arah Sebagai Ekstraksi Fitur," *Jurnal Rekayasa Sistem Informasi*, vol. 4, no. 1, pp. 85–92, 2021.
- [3] A. El-Sawy, M. Loey, and H. El-Bakry, "Arabic handwritten characters recognition using convolutional neural network," *WSEAS Transactions on Computer Research*, vol. 5, no. 1, pp. 11–19, 2017.
- [4] D. Gupta and S. Bag, "CNN-based multilingual handwritten numeral recognition: A fusion-free approach," *Expert Syst Appl*, vol. 165, p. 113784, 2021.
- [5] Z. Deng, X. Zhu, D. Cheng, M. Zong, and S. Zhang, "Efficient kNN classification algorithm for big data," *Neurocomputing*, vol. 195, pp. 143–148, 2016.
- [6] L. Xiong and Y. Yao, "Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm," *Build Environ*, vol. 202, p. 108026, 2021.
- [7] L. Wang, "Research and implementation of machine learning classifier based on KNN," in *IOP Conference Series: Materials Science and Engineering*, IOP publishing, 2019, p. 052038.
- [8] Y. Wang, R. Wang, D. Li, D. Adu-Gyamfi, K. Tian, and Y. Zhu, "Improved handwritten digit recognition using quantum k-nearest neighbor algorithm," *International Journal of Theoretical Physics*, vol. 58, pp. 2331–2340, 2019.
- [9] M. Turmudzi and E. Setyati, "Identifikasi Penulis Berdasarkan Pola Tulisan Tangan Menggunakan Convolutional Autoencoder dan KNN," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 3, no. 1, pp. 8–13, 2021.
- [10] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," *Soft comput*, vol. 25, no. 3, pp. 2277–2293, 2021.