

PENENTUAN KLASTER PADA ANALISIS TULISAN TANGAN UNTUK PENGENALAN PRIBADI MENGGUNAKAN METODE CNN- k -MEANS

Derry Alamsyah^{1*}, dan Samsuryadi²

¹ Informatika, Universitas Multi Data Palembang

² Teknik Informatika, Universitas Sriwijaya

Corresponding author: derry@mdp.ac.id

ABSTRAK: Analisis tulisan tangan merupakan bidang penelitian yang mampu memberikan informasi yang berguna untuk berbagai bidang baik psikologi, kesehatan, ilmu komputer dan sebagainya. Pengelompokan atau penentuan klaster pada analisis tulisan tangan menjadi sangat penting untuk kondisi data yang tidak memiliki label, terutama label sifat atau karakter (pribadi). Hal ini dapat menjadi penelitian dasar untuk menentukan pola tulisan tangan dalam mengenali sifat karakter (pribadi) seseorang. Oleh karena itu, penelitian ini menggunakan k -Means sebagai metode klaster yang memiliki kemudahan dan performa yang baik dalam klasterisasi. Sementara itu, ekstraksi fitur yang digunakan pada penelitian ini adalah CNN. Jumlah klaster yang digunakan pada penelitian ini adalah 2, 3, 4 dan 5 bersesuaian dengan skala Five-Factor Model (FFM) dalam menentukan pribadi seseorang. Dataset yang digunakan pada penelitian ini adalah *public dataset* yaitu *IAM-online Handwritten Dataset*. Hasil penelitian menunjukkan adanya pola klaster yang menyebar (*sparse*) dan saling tumpang tindih. Hal tersebut ditunjukkan oleh nilai *silhouette coefficient* yang mendekati nol.

Kata Kunci: Analisis, Tulisan, CNN, k -means, Klaster

ABSTRACT: *Handwriting analysis is an area of research that is able to provide useful information for various fields including psychology, health, computer science and so on. Grouping or determining clusters in handwriting analysis becomes essential for the underlying data without annotation or label, in particular labels of personality. This can be a basic research to determine handwriting patterns in recognizing a person's character traits (personality). Therefore, this study uses k -Means as a cluster method that has ease and good performance in clustering. Meanwhile, the feature extraction used in this study is CNN. The number of clusters used in this study were 2, 3, 4 and 5 according to the Five-Factor Model (FFM) scale in determining a person's personality. The dataset used in this study is a public dataset, namely the IAM-online Handwritten Dataset. The results showed that there was a sparse and overlapping cluster pattern. This is indicated by the silhouette coefficient value which is close to zero.*

Keyword: *Analysis, Handwritten, CNN, k -means, cluster*

PENDAHULUAN

Analisis tulisan tangan (*handwritten analysis*) merupakan cara atau pendekatan untuk mendapatkan informasi tertentu dari tulisan tangan seseorang. Informasi yang didapatkan dapat berupa kondisi psikologi, sifat atau karakter (pribadi), atribut fisik seperti jenis kelamin, usia dsb., atau bahkan deteksi suatu penyakit. Penentuan atau deteksi jenis penyakit telah dilakukan oleh Bhaskoro dan Supangkat (2014), yaitu melakukan deteksi penyakit diabetes melalui analisis tulisan tangan. Dari penelitiannya akurasi didapat akurasi diatas 80%. Performa yang sama juga dimiliki oleh Drotar et. al. (2013) sebelumnya. Pada penelitiannya, mereka

melakukan deteksi micrographia yang dilakukan untuk mengenali gejala penyakit Parkinson.

Dalam penentuan atribut fisik, beberapa penelitian telah dilakukan diantaranya Siddiqi et. al. (2015) menggunakan metode Artificial Neural Network (ANN) dan Support Vector Machine untuk mengenali jenis kelamin seseorang melalui tulisan tangan. Dalam penelitiannya fitur yang digunakan adalah kemiringan (*slant*), lengkungan (*curvature*), tekstur (*texture*) dan keterbacaan (*legibility*). Hasilnya, akurasi sebesar 80% sebagai performa sistem yang ditunjukkan. Kemudian, ada juga Maaded dan Hassaine (2014) sebelumnya, melakukan hal serupa, termasuk mengenali atribut lainnya seperti usia dan kebangsaan. Berbeda dengan Siddiqi et.

al, pada penelitiannya Maaded menggunakan fitur geometri. Hasilnya dibagi kedalam dua jenis pengujian, yaitu pengujian terhadap skenario penulisan teks yang sama yaitu 75% untuk usia, 55,67% untuk jenis kelamin dan 53,66% untuk kebangsaan. Sementara itu 73,59%, 60,62% dan 47,98% merupakan hasil untuk atribut fisik yang sama terhadap skenario penulisan teks yang berbeda.

Selain itu semua, analisis tulisan tangan juga dapat digunakan untuk menentukan cara penulisan anak-anak yang nantinya dapat digunakan oleh ahli analisis psikologi untuk menentukan cara meningkatkan kualitas pendidikan anak (Zaarour et. al., 2004). Data yang digunakan pada penelitian ini berupa gambar yang ditulis/lukis oleh anak-anak dan menggunakan pendekatan Bayesian-Network (BN) model. Kemudian, gejala gangguan perkembangan anak dapat dideteksi melalui tulisan tangannya. Penelitian tersebut dilakukan oleh Sudirman et. al. (2011), dimana hasil akurasi yang didapatkan adalah sebesar 78%.

Umumnya analisis tulisan tangan dilakukan untuk menentukan atau mengidentifikasi sifat atau karakter (pribadi) seseorang. Champa dan Anandakumar (2010) menggunakan ciri/fitur huruf "t" kecil, lingkaran pada bagian bawah huruf "y", tekanan pena, kemiringan penulisan (slant of writing). Kemudian menggunakan pendekatan klasifikasi berbasis aturan menggunakan skala Myers-Briggs Type Indicator (MBTI). Hasil yang didapat adalah akurasi sebesar 70%. Kemudian, Benham Fallah dan Hassan (2016) Khotanlou melakukan penelitian analisis tulisan tangan menggunakan dataset Minnesota Multiphasic Personality Inventory (MMPI). Mereka menggunakan NN untuk mengenali atribut yang tidak berkaitan dengan penulis, sementara HMM digunakan sebaliknya. Hasil akurasi yang dimiliki oleh penelitian ini adalah 70%. Chen dan Tao (2017) selanjutnya menggunakan SVM, AdaBoost, dan kNN untuk mengklasifikasi setiap tujuh sifat dengan tujuan menganalisis sejumlah ciri khusus. Mereka menghasilkan akurasi mulai dari 62.5 s/d 83.9%. Selain itu, Gorbova et. al. (2017) mengembangkan sistem yang mampu memberikan skor untuk skala FFM melalui isyarat visual, audio dan lexical. Hasil yang diberikan cukup menjanjikan dalam upaya untuk membantu seleksi pelamar pekerjaan.

Dalam melakukan analisis tulisan tangan untuk mengenali sifat atau karakter seseorang terdapat dua teori yang dapat dijadikan alat untuk mencapai, diantaranya skala Five-Factor Model (FFM) dan analisis Graphologi. FFM merupakan sebuah model untuk menggambarkan sifat dari seseorang. Hal ini didasarkan pada lima sifat dasar yaitu: *Openness to Experience*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism*.

FFM sukses digunakan pada berbagai macam peran penugasan. Riset yang dilakukan oleh Roberts et. al. pada tahun 2008 menunjukkan bahwa dibandingkan model lain, FFM memberikan kestabilan relatif terhadap waktu yang cukup lama (4 tahun). FFM juga menunjukkan dapat digunakan untuk menentukan kelainan sifat/pribadi

seperti depresi, kecemasan dan bahkan penyakit fisik seperti masalah hati, kanker, diabetes atau permasalahan pernapasan. Selain itu juga sukses digunakan pada wilayah pengembangan dan konsultasi karir serta performa tim dan juga peningkatan cara belajar dan performa akademis siswa.

Sementara itu, ketika menganalisis tulisan tangan seseorang (analisis graphologi), seorang ahli graphologi akan memperhatikan ciri khusus diantaranya garis dasar (*baseline*), kemiringan huruf (*word slant*), penekanan tulisan (*writing pressure*), tulisan sambung (*connecting strokes*), jarak antar baris (*space between lines*), huruf "t" kecil (*lowercase letter of "t"*), dan huruf "f" kecil (*lowercase letter "f"*).

Tidak terdapat standar dalam mengembangkan pengenalan sifat/pribadi berdasarkan tulisan tangan. Secara umum analisis graphologi dilakukan oleh para ahli graphologi (graphologist). Akan tetapi, pada bidang ilmu komputer penelitian dilakukan untuk membangun atau mengembangkan sistem yang mampu mengenali sifat melalui tulisan tangan dengan cara yang lebih mudah serta membakukan standar analisis graphologi.

Oleh karena itu penelitian ini merupakan penelitian dasar yang dilakukan untuk mengetahui pola kelompok atau klaster dari tulisan tangan seseorang. Sehingga, nantinya dapat menjadi acuan atau dasar informasi untuk menentukan kelas atau kelompok sifat atau karakter (pribadi) dasar seseorang seperti yang dimiliki oleh skala FMM. Dalam penelitian ini, skema klaser yang diajukan menggunakan ekstraksi fitur dengan metode CNN dan klasterisasi dengan metode *k-means*.

CONVOLUTIONAL NEURAL NETWORK (CNN)

Pendekatan dalam klasifikasi yang populer saat ini adalah deep learning, dimana pendekatan ini mampu merekayasa fitur atau ciri secara otomatis (automated feature engineering) dan memberikan akurasi pengenalan yang tinggi pada umumnya. Salah satu metode dalam pendekatan ini adalah Convolutional Neural Network (CNN). Metode CNN merupakan pendekatan deep learning yang banyak digunakan untuk atau diaplikasikan pada analisis citra. CNN serupa dengan metode Artificial Neural Network (ANN) kecuali pada bagian konvolusinya. Kesamaan atau hal serupa tersebut ditunjukkan oleh Gambar 1 pada bagian fully-connected-nya. Metode ini terdiri dari beberapa layer dalam arsitekturnya (ditunjukkan oleh Gambar 1), diantaranya convolutional layer, pooling layer, normalization layer, ReLu layer, fully-connected layer dan loss layer (Suyanto, 2018).

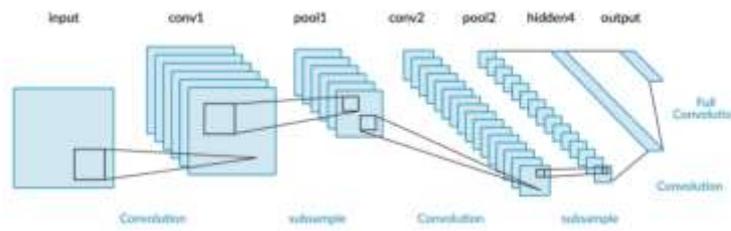
1. Convolutional Layer

Layer ini merupakan layer yang mencari ciri khas dari metode CNN. Pada layer ini citra masukkan diproses dengan menggunakan metode konvolusi. Tujuan layer ini adalah untuk menghasilkan ciri dari suatu citra

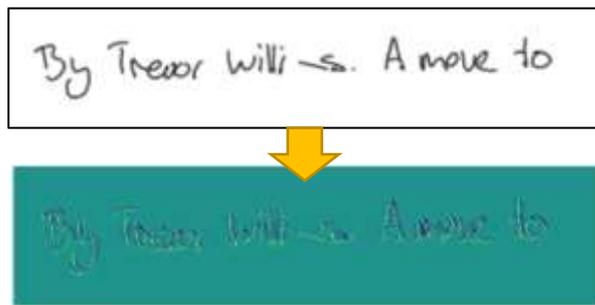
(ditunjukkan oleh Gambar 2). Konvolusi yang dilakukan pada metode ini sama halnya konvolusi yang dilakukan pada metode pengolahan citra umumnya. Proses konvolusi dilakukan secara berurutan dengan melibatkan kernel yang dilakukan pada setiap sub-citra (ditunjukkan oleh Gambar 3 a: sub-citra ditunjukkan dengan area berwarna biru). Setiap proses sekuensialnya menghasilkan satu luaran baru (titik merah Gambar 3.a). Sementara sebagian ciri yang dihasilkan dari sub-citra ditunjukkan oleh Gambar 3 b.

2. Pooling Layer

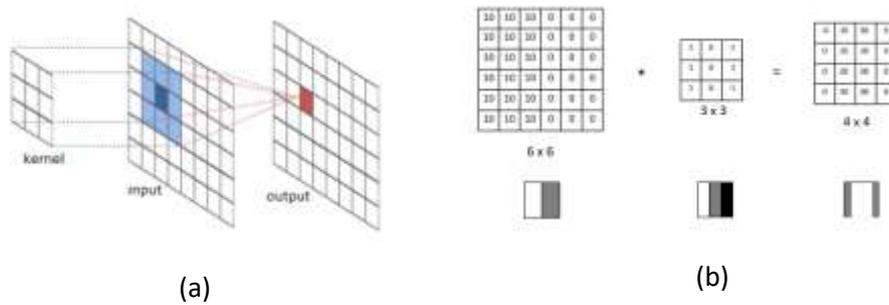
Layer ini berfungsi untuk mereduksi ukuran (down sampling) dari citra masukan sehingga menghasilkan ukuran citra yang lebih kecil untuk proses selanjutnya. Proses reduksi pada citra dapat dilakukan dengan cara menentukan nilai maksimum dari suatu wilayah sub-citra. Hal tersebut disebut sebagai max pooling (ditunjukkan oleh Gambar 4). Selain menggunakan metode max pooling, masih terdapat metode lain salah satunya average pooling yaitu dengan mencari nilai rata-rata dari suatu wilayah sub-citra.



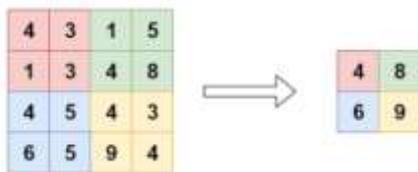
Gambar 1 Arsitektur Convolutional Neural Network (CNN)



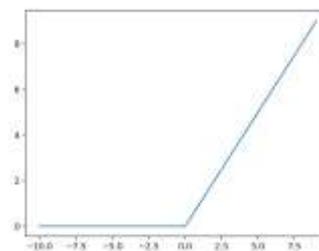
Gambar 2 Feature engineering



Gambar 3 (a) Proses Konvolusi dan (b) Hasil Konvolusi



Gambar 4 Poolng Layer



Gambar 5 Grafik fungsi ReLu

3. Pooling Layer

Nilai intensitas citra masukan dapat sangat bervariasi, oleh karena itu untuk mengatasi perbedaan rentang nilai tersebut dilakukan proses normalisasi. Ada banyak cara untuk melakukan normalisasi dalam pengolahan citra, yang paling umum adalah standar normal rata-rata nol dan normalisasi rentang [0, 1].

$$N(r, c) = I(r, c)/255 \quad (1)$$

Dalam metode CNN, metode normalisasi yang sering dilakukan adalah normalisasi rentang [0, 1], yang artinya nilai intensitas pada citra yang dihasilkan diubah kedalam skala pada rentang [0, 1].

Metode ini cukup sederhana yaitu dilakukan dengan membandingkan nilai saat ini (pada pixel posisi tertentu) dengan nilai intensitas maksimum (ditunjukkan oleh persamaan 1). Nilai intensitas maksimum pada citra 8-bit adalah 255. Akan tetapi, karena pengaruhnya yang relatif kecil pada performa metode CNN secara keseluruhan, maka umumnya layer ini ditinggalkan.

4. ReLu Layer

Layer ini bertujuan untuk mengimplementasikan fungsi aktivasi, khusus fungsi (2). Fungsi ini meningkatkan sifat nonlinearitas fungsi keputusan dan jaringan (network) secara keseluruhan. Fungsi ini merupakan fungsi sederhana yang mengubah semua nilai dibawah nol (negatif) menjadi nol. Grafik dari fungsi ini ditunjukkan oleh Gambar 5, sementara dapat dari fungsi ini ditunjukkan oleh Gambar 2. Selain menggunakan fungsi (2), CNN juga dapat menggunakan berbagai fungsi aktivasi lainnya seperti tangen hiperbolik (3) atau fungsi eksponen (4).

$$f(x) = \max(0, x) \quad (2)$$

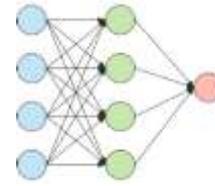
$$f(x) = \tanh(x) \quad (3)$$

$$f(x) = (1 + e^{-x})^{-1} \quad (4)$$

5. Fully Connected Layer

Layer ini merupakan layer dengan arsitektur yang sama halnya dengan milik metode ANN. Tujuan pada layer ini pun serupa, yaitu untuk memberikan bobot pada fitur yang diberikan yang nantinya dapat digunakan untuk mengenali kelas atau kelompok dari citra masukan.

Untuk CNN yang digunakan sebagai ekstraksi ciri, layer ini dapat ditinggalkan. Penentuan bobot pada layer ini dapat dilakukan dengan menggunakan algoritma backpropagation, sementara untuk pada setiap layer yang terhubung di layer ini dihubungkan dengan menggunakan fungsi aktivasi. Contoh susunan atau arsitektur pada layer ini ditunjukkan oleh Gambar 6 yang digunakan untuk memprediksi dua kelas.



Gambar 6 Fully Connected Layer

6. Loss Layer

Merupakan layer terakhir yang bertujuan untuk menentukan penalti dari hasil yang tidak bersesuaian dengan data ground truth (data label/acuan/target) pada proses pelatihan (penentuan bobot pada fully connected layer). Beberapa fungsi yang dapat digunakan pada layer ini diantaranya adalah sigmoid cross-entropy loss, softmax loss, euclidean loss dsb.

K-MEANS

Algoritma atau metode *k*-means merupakan algoritma yang populer dalam melakukan klusterisasi. Hal tersebut dikarenakan kemudahan dalam pengimplementasiannya dan kecepatan dalam pemrosesannya (Suyanto, 2018). Contoh penerapan metode ini pada ruang dimensi dua ditunjukkan oleh Gambar 7.

Selanjutnya, diberikan suatu himpunan data yang berisikan *n*-objek $D = \{d_1, d_2, \dots, d_n\}$ dimana $d_i \in R^n$. Kemudian akan dicari kluster (label) untuk setiap objek ke-*i* sehingga terdapat himpunan pasangan data dan kluster yaitu (d_i, l_i) , dimana $l_i = \{1, 2, \dots, c\}$ dengan *c* sebagai indeks kluster (label) merupakan tujuan dari metode *k*-means. Kluster-kluster tersebut dapat dinyatakan sebagai $C = \{C_1, C_2, \dots, C_k\}$ dengan *k* adalah jumlah kluster (sesuai dengan nama dari metode ini). Penentuan kluster l_i untuk setiap data d_i dilakukan dengan menentukan nilai minimum sum of within-cluster scatters (5):

$$l_i = \underset{c}{\operatorname{argmin}} \sum_{j=1}^k \sum_{d_i \in C_j} \operatorname{dist}(d_i, \mu_j) \quad (5)$$

Dimana μ_j merupakan rata-rata (centroid) dari kelas $C_j \in C$. Penentuan nilai μ_j dilakukan dengan cara (6):

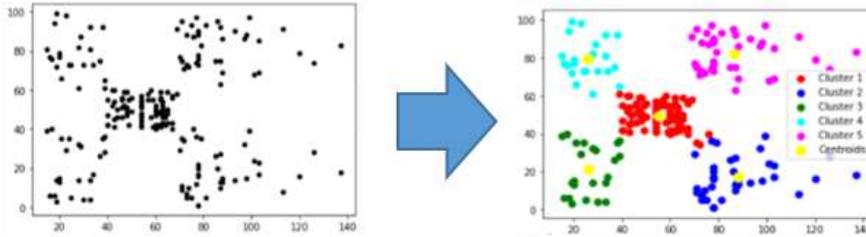
$$\mu_j = |C_j|^{-1} \sum_{d_i \in C_j} d_i \quad (6)$$

Algoritma atau metode ini bertindak secara inkremental sehingga nilai dari μ_j untuk setiap $j = \{1, \dots, k\}$ konvergen. Atau dengan kata lain $\left\| \mu_j^t - \mu_j^{t-1} \right\|_{\infty} < \epsilon$.

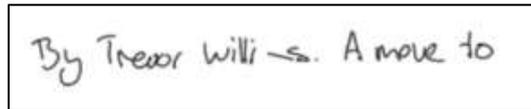
SILHOUETTE COEFFICIENT

Evaluasi atau penilaian performa dari suatu metode klusterisasi dapat dilakukan dengan menggunakan penilaian tendensi klusterisasi, penentuan jumlah kluster dan pengukuran kualitas kluster. Pengukuran kualitas kluster dapat dilakukan dengan dua metode, diantaranya metode ekstrinsik dan intrinsik.

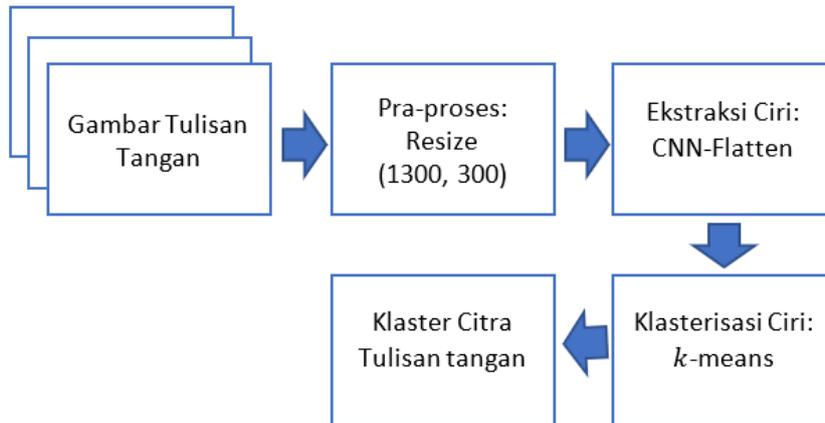
Metode ekstrinsik dapat dilakukan jika terdapat kluster dari populasi atau kluster acuan, yaitu dengan cara membandingkan hasil kluster yang diperoleh dari metode klusterisasi tertentu dengan kluster acuan yang kita punya tersebut. Sementara, metode intrinsik merupakan metode pengukuran kualitas kluster yang bertolak belakang dengan metode ekstrinsik. Pada metode ini, kita tidak memerlukan kluster acuan yang tentunya merupakan kondisi yang paling sering kita dapati.



Gambar 7. Algoritma k-means.



Gambar 8. Contoh Gambar Tulisan Tangan



Gambar 9. Rancangan Sistem Klusterisasi

Diberikan suatu himpunan data yang berisikan n -objek $D = \{d_1, d_2, \dots, d_n\}$. Selanjutnya, terdapat $C_j \subseteq D$ yang merupakan kluster (himpunan bagian) dari data D . Maka dapat diukur (Suyanto, 2018):

1. Kepadatan Kluster ($a(d)$)

$a(d_i)$ (persamaan 7) dimana $d_i \in C_k$ merupakan rata-rata jarak antara objek d_i dengan $\forall d_j \in C_k$ dimana $i \neq j$. Nilai $a(d_i)$ menyatakan kepadatan kluster yang mengandung objek d_i . Semakin kecil $a(d_i)$, semakin padat kluster tersebut.

$$a(d_i) = \frac{\sum_{d_j \in C_k, i \neq j} dist(d_i, d_j)}{|C_k| - 1} \tag{7}$$

2. Jarak Objek dan Kluster ($b(d)$)

$b(d_i)$ (persamaan 8) dimana $d_i \in C_l$ merupakan rata-rata minimum dari jarak objek d_i dengan C_k dimana $k \neq l$. Nilai $b(d_i)$ menggambarkan seberapa jauh objek d_i terpisah dari kluster-kluster lain. Semakin besar nilai $b(d_i)$ maka semakin jauh objek d_i dengan kluster-kluster lain.

$$b(d_i) = \min_{C_k, k \neq l} \left\{ \frac{\sum_{d_j \in C_k} dist(d_i, d_j)}{|C_k|} \right\}, d_i \in C_l \tag{8}$$

3. Silhouette Coefficient ($s(d)$)

Jika $a(d_i)$ bernilai sangat kecil, sementara $b(d_i)$ bernilai sangat besar maka nilai silhouette coefficient $s(d_i)$ akan mendekati 1. Hal ini berarti kluster yang berisi

objek d_i sangat padat (dense) dan objek d_i itu sendiri terpisah jauh dari kluster-kluster lainnya.

$$s(d_i) = \frac{b(d_i) - a(d_i)}{\max\{a(d_i), b(d_i)\}} \quad (9)$$

Jika $a(d_i)$ bernilai sangat besar, sementara $b(d_i)$ bernilai sangat kecil maka nilai silhouette coefficient $s(d_i)$ pada persamaan 9 akan mendekati -1. Hal ini berarti kluster yang berisi objek d_i tidak padat (sparse) dan objek d_i berada sangat dekat atau bahkan tumpang tindih dengan kluster lainnya.

Pengukuran kualitas suatu kluster dalam suatu proses klusterisasi dilakukan dengan cara menghitung rata-rata nilai silhouette coefficient s dari semua objek d dalam suatu kluster C . Sementara kualitas klusterisasi dari algoritma atau metode tertentu diukur dengan menghitung rata-rata nilai silhouette coefficient s dari semua objek d pada himpunan data D .

METODOLOGI PENELITIAN

Pada penelitian ini metodologi yang digunakan terdiri dari tiga cara/tahap yaitu pengumpulan data, rancangan sistem dan evaluasi.

1. Pengumpulan Data

IAM-Online merupakan dataset tulisan tangan yang diambil melalui penulisan langsung dengan perangkat (bukan merupakan hasil scan tulisan tangan). Dataset ini terdiri dari 146 dokumen dan 195 penulis. Ukuran citra yang ada pada dataset ini beragam namun berupa persegi, mengikuti pola penulisan kalimat. Tiap dokumen yang ditulis oleh penulis dibagi kedalam delapan citra hasil. Citra yang dihasilkan menggunakan format TIFF. Contoh citra tulisan tangan IAM-Online Dataset ditunjukkan oleh Gambar 8. Dalam penelitian ini digunakan 10 dokumen

yang diambil secara acak. Dari 10 dokumen yang terpilih terdapat 119 penulis. Pengurangan jumlah sampel ditujukan untuk memenuhi kebutuhan perangkat yang ada.

2. Rancangan Sistem

Sistem klusterisasi yang dibangun ditunjukkan oleh Gambar 9. Pada sistem ini citra tulisan tangan yang berasal dari IAM-online dataset dilakukan perubahan ulang ukuran yaitu menjadi (1300, 200). Hal tersebut bertujuan untuk menstandarkan ukuran citra masukan. Selanjutnya, dengan menggunakan hasil flatten dari metode CNN maka didapat ciri dari citra masukan. Kemudian dengan menggunakan metode k -means maka dilakukan klusterisasi yang dibagi kedalam 4 kelompok pengujian yaitu kluster dengan jumlah 2, 3, 4, dan 5..

3. Evaluasi

Proses penentuan nilai performa dari metode klusterisasi yang digunakan dalam penelitian ini adalah pengukuran kualitas kluster menggunakan metode intrinsik yaitu silhouette coefficient.

PEMBAHASAN

Pada penelitian digunakan arsitektur CNN dengan tiga lapisan konvolusi (*convolution layer*), dengan masing-masing satu lapisan reduksi fitur (*max pooling layer*) dan lapisan vektor fitur (*flatten*) yang ditunjukkan oleh Gambar 10. Tiap lapisan konvolusi memiliki jumlah parameternya masing-masing, dimana parameter ini terdiri dari dua jenis parameter yaitu bobot (*weight*) dan bias. Dalam hal ini, parameter bobot dihasilkan dari operasi konvolusi. Selanjutnya jumlah masing-masing parameter lapisan konvolusi dapat dihitung dengan menggunakan persamaan 10.

$$\text{Jumlah param. laps. konvolusi} = ((\text{Ukuran Kernel} \times \text{kanal layer sebelumnya}) + 1) \times \text{kanal layer saat ini} \quad (10)$$

Layer (type)	Output Shape	Param #
conv2d_6 (Conv2D)	(None, 296, 1196, 32)	832
max_pooling2d_6 (MaxPooling2)	(None, 148, 598, 32)	0
conv2d_7 (Conv2D)	(None, 146, 596, 32)	9248
max_pooling2d_7 (MaxPooling2)	(None, 73, 298, 32)	0
conv2d_8 (Conv2D)	(None, 71, 296, 32)	9248
max_pooling2d_8 (MaxPooling2)	(None, 35, 148, 32)	0
flatten_2 (Flatten)	(None, 165760)	0
Total params: 19,328		
Trainable params: 19,328		
Non-trainable params: 0		

Gambar 10. Arsitektur CNN yang digunakan

Perhitungan untuk jumlah masing-masing parameter dilakukan sebagai berikut:

Jumlah parameter lapisan konvolusi pertama :

$$((5^2 \times 1) + 1) \times 32 = 832$$

Jumlah parameter lapisan konvolusi kedua :

$$((3^2 \times 32) + 1) \times 32 = 9248$$

Jumlah parameter lapisan konvolusi ketiga :

$$((3^2 \times 32) + 1) \times 32 = 9248$$

Sementara itu, lapisan reduksi fitur tidak memiliki parameter. Selanjutnya, Kernel untuk lapisan konvolusi pertama ditunjukkan oleh gambar 11 dengan ukuran dari masing-masing kernel adalah 5×5 . Selanjutnya untuk setiap kernel yang diberikan, dilakukan proses konvolusi sehingga menghasilkan matriks ciri (*feature map*) sebagai hasil dari setiap layer ini. Matriks ciri untuk tiap lapisan konvolusi ditunjukkan oleh Tabel 1.

Tabel 1 Ekstraksi Fitur per Layer

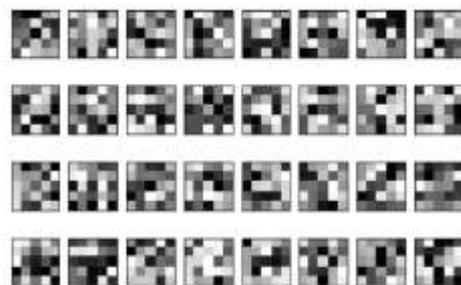
Lapisan (Layer)	Hasil
Convolution 2d (296, 1192, 32)	
Convolution 2d (146, 596, 32)	
Convolution 2d (71, 296, 32)	
Skala Warna (viridis)	

Matrix ciri terakhir dari proses konvolusi (yang telah direduksi) dijadikan vektor ciri. Hal ini dapat dilakukan dengan melakukan transformasi bentuk dari bentuk matrix ke bentuk vektor atau dikenal dengan istilah *flatten*. Dalam penelitian ini CNN difungsikan menjadi metode ekstraksi ciri (*feature extractor*). Ciri yang dihasilkan oleh arsitektur CNN dalam penelitian ini berjumlah 19,328 ciri per citra.

Selanjutnya, dilakukan proses klusterisasi (*clustering*) untuk setiap titik data $d \in R^{19,328}$ dengan menggunakan metode *k*-means. Pengukuran kemiripan antar titik data yang dilakukan untuk menentukan nilai *centroid* dilakukan menggunakan pengukuran jarak Euclidean. Penentuan jumlah *k* didasari oleh lima skala dari *Five-Factor Model* (FFM), sehingga jumlah kluster yang dipilih berada pada rentang [2, 5]. Dimensi yang besar dari setiap titik data *d* membuat kita tidak dapat melakukan pengamatan kualitas kluster secara spasial, oleh karenanya digunakan metode *silhouette coefficient* untuk menentukan kualitas dari kluster yang terbentuk.

Pengukuran nilai *silhouette coefficient* dilakukan untuk masing-masing titik data yaitu $s(d)$. Selanjutnya kualitas dari kluster ditunjukkan oleh nilai rata-rata dari $s(d)$. Pada setiap kluster yang terbentuk, baik 2, 3, 4 ataupun 5 seluruhnya menghasilkan nilai rata-rata $s(d)$ yang mendekati nol (perhatikan Tabel 2). Hal tersebut menunjukkan hampir imbangnya jumlah $s(d)$ yang bernilai positif dan negatif pada setiap *d* dalam kluster

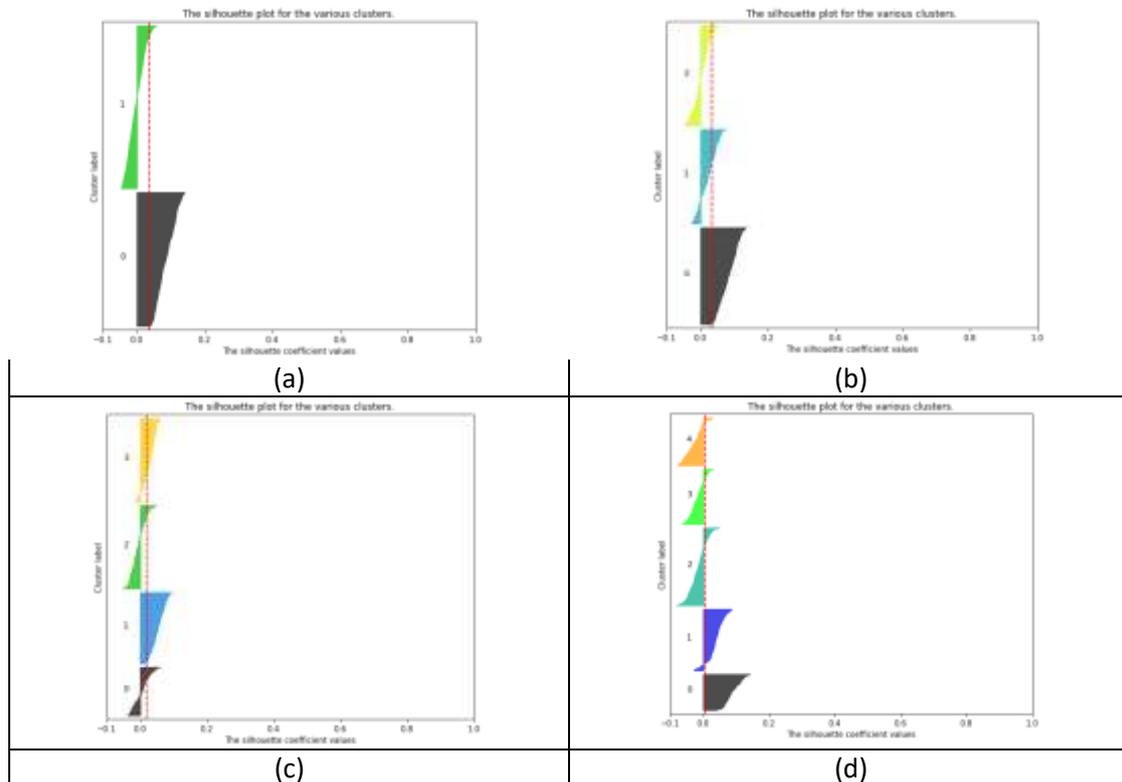
tertentu C_k . Seperti yang telah dijelaskan pada bagian (*section*) *silhouette coefficient* nilai $s(d)$ yang bernilai positif menunjukkan adanya kecenderungan kluster yang padat dan berjauhan, begitupun sebaliknya untuk nilai $s(d)$ yang negatif. Berdasarkan Gambar 12, kita dapat melihat bahwa distribusi $s(d)$ berada pada rentang yang cukup jauh dari nilai 1 ataupun -1. Hal ini mengindikasikan bahwa nilai $a(d)$ sebagai jarak intrakluster dan $b(d)$ sebagai jarak interkluster tidak terlalu berbeda, dengan kata lain pola kluster yang terbentuk menumpuk dalam satu kluster utama dan cukup padat.



Gambar 10 Kernel pada Lapisan Konvolusi Pertama

Tabel 2. Hasil $S(d)$

Jumlah Kluster	Hasil Rata-rata $s(d)$
2	0.037
3	0.035
4	0.02
5	0.006



Gambar 12 Hasil Evaluasi

KESIMPULAN

Pada penelitian ini dapat disimpulkan bahwa:

1. Kualitas klaster yang dihasilkan tidak cukup baik, hal ini dikarenakan nilai *silhouette coefficient* yang mendekati nol.
2. Kualitas klaster yang tidak baik ini menunjukkan setiap titik data yang merupakan ciri tulisan tangan yang dihasilkan oleh metode CNN membentuk sebaran yang cukup rapat dan saling tumpang tindih.

SARAN

Penelitian selanjutnya diharapkan menggunakan variasi dari arsitektur CNN dan menggunakan jumlah data yang lebih banyak lagi. Penggunaan data yang lebih banyak tentunya akan membutuhkan dukungan perangkat yang bersesuaian. Akan tetapi hal ini dapat diatasi dengan variasi pengambilan data dalam menentukan *centroid* dari metod *k-means*

DAFTAR PUSTAKA

B Fallah, H Khotanlou, in Artificial Intelligence and Robotics (IRANOPEN). Identify human personality parameters based on handwriting using neural networks (April 2016)
 BW Roberts, D Mroczek, Personality trait change in adulthood. *Curr. Dir. Psychol. Sci.* 17(1), 31–35 (2008)
 Drotar, P., Mekyska, J., Smekal, Z., Rektorova, I., Prediction potential of different handwriting tasks for

diagnosis of Parkinson's, 2013 E-Health and Bioengineering Conference, Pages 1–4, November 2013.

HN Champa, KR Anandakumar, Automated human behavior prediction through handwriting analysis. 2010 First International Conference on Integrated Intelligent Computing (ICIIC), 160–165 (August 2010)

I Zaarour, L Heutte, P Leray, J Labiche, B Eter, D Mellier, Clustering and Bayesian network approaches for discovering handwriting strategies of primary school children. *Int. J. Pattern Recognit. Artif. Intell.* 18(7), 1233–1251 (2004)

J Gorbova, I Lusi, A Litvin, G Anbarjafari, Automated screening of job candidate based on multimodal video processing. *Computer Vision and Pattern Recognition Workshops (CVPRW) (2017) IEEE Conference on*, July 2017

R Sudirman, N Tabatabaey-Mashadi, I Ariffin, Aspects of a standardized automated system for screening children's handwriting. *First international conference on Informatics and Computational Intelligence (ICI)*, 48–54 (December 2011)

SB Bhaskoro, SH Supangkat, An extraction of medical information based on human handwritings. 2014 International Conference on Information Technology Systems and Innovation (ICITSI), 253–258 (November 2014)

Suyanto, "Machine Learning Tingkat Dasar dan Lanjut", 2018, Informatika.

Z Chen, T Lin, Automatic personality identification using writing behaviors: an exploratory study. *Behav Inform Technol* 36(8), 839–845 (2017)