

ANALISIS PENGUKURAN NUMERIK PADA OPTIMALISASI CLUSTER OBAT DI APOTEK X MENGGUNAKAN DAVIES BOULDIN INDEX

Gusrianty¹, Muhammad Siddik²

^{1,2}Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia, Jl. Jend. Ahmad Yani No. 78-88
Pekanbaru, Riau

email: ¹gusrianty@lecturer.pelitaindonesia.ac.id, ²siddik@lecturer.pelitaindonesia.ac.id

Abstract

The use of different measuring techniques in grouping results in different grouping results. Testing the data on a number of k tests makes the placement of drug cluster members not the same in each measurement so that cluster optimization is then needed. This can raise doubts for data users who want to obtain accurate information. The study was conducted to examine the comparison of drug cluster optimization results using category, type and drug unit variables grouped by numerical measurements namely Manhattan Distance, Canberra Distance and Dynamic Time Warping Distance (DTWD) using the Davises Bouldin Index (DBI) cluster evaluation technique. The test results obtained through the RapidMiner application show that the optimal cluster of the three measurements is at k=2, k=3 and k=2 with DBI values of 0.752, 0.873, and 0.868. Manhattan was chosen as the better numerical measurement technique. than the other two techniques because it has the lowest DBI value tested at k 2.

Keywords: *K test, cluster optimalization, numerical measurements, DBI*

Abstrak

Penggunaan teknik ukur yang berbeda dalam pengelompokkan mengakibatkan hasil pengelompokkan yang berbeda pula. Pengujian data terhadap sejumlah k uji menjadikan penempatan anggota cluster obat tidak sama pada setiap pengukuran sehingga kemudian diperlukan optimalisasi cluster. Hal ini dapat menimbulkan keraguan bagi para pengguna data yang ingin memperoleh informasi yang akurat. Penelitian dilakukan untuk mengkaji perbandingan hasil optimalisasi cluster obat dengan menggunakan variabel ketegori, jenis dan satuan obat yang dikelompokkan dengan pengukuran numerik yakni Manhattan Distance, Canberra Distance dan Dynamic Time Warping Distance (DTWD) menggunakan teknik evaluasi cluster Davises Bouldin Index (DBI). Hasil pengujian yang diperoleh melalui aplikasi RapidMiner menunjukkan bahwa cluster optimal dari ketiga pengukuran tersebut terdapat pada k=2, k=3 dan k=2 dengan nilai DBI adalah 0,752, 0,873, dan 0, 868. Manhattan terpilih sebagai teknik ukur numerik yang lebih baik dari pada dua teknik lainnya karena memiliki nilai DBI terendah yang di uji pada k 2.

Keywords: *K uji, Optimalisasi, Cluster, Pengukuran numerik, DBI*

1. PENDAHULUAN

Hasil pengelompokkan (*cluster*) terbaik merupakan tujuan melakukan optimalisasi *cluster*. *Cluster* membentuk kumpulan data berdasarkan karakteristik alaminya (Prasetyo, 2014) namun tidak memerlukan target luaran (Nishom & Fathoni, 2018) sehingga pengelompokkan yang terjadi menimbulkan keraguan.

Pengelompokkan dapat dilakukan dengan berbagai teknik pengukuran numerik. Proses yang berbeda pada setiap pengukuran menghasilkan analisis yang berbeda. Saat dilakukan optimalisasi terhadap *cluster* obat, besar kemungkinan *cluster* optimal berada pada k yang berbeda. Pengguna data selalu dihadapkan pada permasalahan menentukan teknik pengukuran numerik yang sesuai dengan data. Penelitian ini bertujuan untuk memberikan informasi tentang *variabel* data uji seperti ketegori, jenis dan satuan obat.

Pengukuran numerik, menggunakan metode *Manhattan Distance*, *Canberra Distance* dan *Dynamic Time Warping Distance* dalam membentuk *cluster* obat yang optimal menggunakan teknik evaluasi *cluster Davies Bouldin Index* (DBI). Dalam penelitian ini tidak melakukan *clusterisasi* pemasaran obat yang laris terjual melainkan menentukan nilai numerik terendah dari ketiga *variabel* yang di uji.

Berbagai penelitian terkait ketiga teknik pengukuran telah banyak dilakukan. Penelitian (Gautama et al., 2015) menunjukkan kemampuan *Manhattan* dalam menganalisis penyimpangan (*anomaly*) trafik saat mengakses internet. *Manhattan* juga digunakan untuk mencari kesamaan bentuk kesalahan yang terjadi pada peralatan mesin (Jiang et al., 2019) dan menentukan posisi karyawan (Miftahuddin et al., 2020). Pada penelitian (Pulungan et al., 2019), pengukuran *Canberra* dapat meningkatkan kinerja algoritma KNN dengan akurasi mencapai 94,70%. Studi yang dilakukan oleh (Putra & Harjoko, 2018), menggunakan *Canberra* untuk pengambilan fitur dari citra wajah dan pengelompokkan jumlah instar pada lalat *Simuliidae* berdasarkan empat kriteria (Yang et al., 2018). Sedangkan untuk teknik pengukuran *Dynamic Time Warping Distance* (DTWD), dapat mengenali kesamaan lisensi nomor kendaraan (Sudarma & Ariyani, 2014) dan pesan *twitter* (Jendoubi et al., 2015).

Ketiga teknik pengukuran yang dijelaskan menggunakan algoritma K-Mean untuk menemukan kesamaan antar data. Sejumlah k diujikan sehingga optimalisasi cluster diperlukan demi mendapatkan kesimpulan terbaik. Studi ini menerapkan teknik *Davies Bouldin Index* (DBI).

Pada buku (Jollyta et al., 2021) dijelaskan bahwa kinerja DBI berdasarkan nilai kohesi dan separasi, dimana dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat cluster dari cluster yang diikuti, sedangkan separasi didasarkan pada jarak antar titik pusat cluster dari cluster. Cluster uji dengan nilai DBI terendah merupakan cluster optimal.

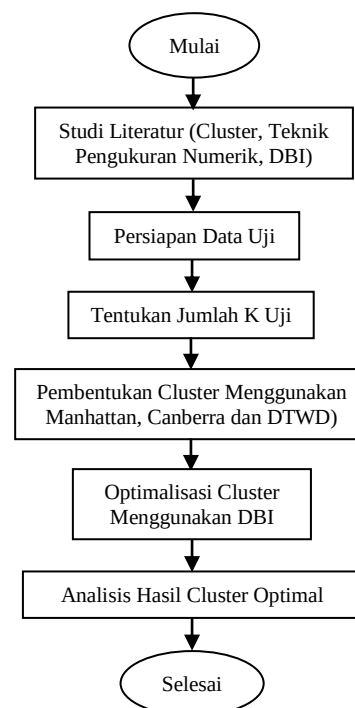
Sejumlah penelitian membuktikan DBI mampu mendapatkan *cluster* optimal,

menggunakan berbagai data seperti data *Covid-19* (Gie & Jollyta, 2020) dan data kecamatan yang dikelompokkan berdasarkan indikator pendidikan (Irhamni et al., 2014). DBI juga berhasil mendapatkan *cluster* optimal melalui tiga macam data set (Sitompul et al., 2019).

Berdasarkan penelitian optimalisasi cluster dengan berbagai teknik pengukuran numerik, dilakukan penelitian ini dengan menggunakan data yang sama yakni data obat dari sebuah Apotek X di Pekanbaru. Adanya perbandingan kinerja ketiga teknik pengukuran dalam menghasilkan *cluster* optimal menggunakan DBI, dapat mengarahkan pengguna pada teknik pengukuran yang tepat sesuai dengan data yang digunakan.

2. METODE PENELITIAN

Untuk memperoleh perbandingan kinerja ketiga teknik pengukuran numerik, disusun tahapan penelitian sebagai berikut:



Gambar 1. Tahapan penelitian

Pada Gambar 1, tahapan penelitian diawali dengan studi literatur tentang ketiga pengukuran numerik dan teknik evaluasi *cluster* DBI. Penjelasan untuk ketiga pengukuran numerik adalah sebagai berikut:

1. Manhattan Distance

Manhattan Distance adalah pengukuran yang dihasilkan berdasarkan jumlah selisih antara dua benda dan hasil yang diperoleh adalah mutlak (Jendoubi et al., 2015). Pengukuran Manhattan menggunakan persamaan (1).

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}| \quad (1)$$

Dimana:

$d(i, j)$ = Jarak data ke i ke pusat cluster j

x_{ni} = Data ke i pada atribut data ke n

x_{nj} = Titik pusat ke j pada atribut ke n

2. Canberra Distance

Canberra Distance bekerja dengan membagi perbedaan mutlak antara variabel dari dua objek dengan jumlah nilai variabel absolut (Lance & Williams, 2012). Pengujian dengan *Canberra* menggunakan persamaan (2).

$$d^{CAD}(i, j) = \sum_{k=0}^{n-1} \frac{|y_{i,k} - y_{j,k}|}{|y_{i,k}| + |y_{j,k}|} \quad (2)$$

3. Dynamic Time Warping Distance (DTWD)

DTWD terkenal karena kemampuannya mengelola distorsi waktu dengan menyelaraskan deret waktu saat membandingkannya (Zhang et al., 2017). Sebagai ukuran kesamaan, DTWD adalah perhitungan jarak terbaik dan lebih mudah dipahami dengan persamaan (3) (Dinov et al., 2016).

$$DTW(Q, C) = \min \left\{ \sqrt{\sum_{k=1}^K W_k} \right\} \quad (3)$$

Selanjutnya dilakukan persiapan data uji. Data yang digunakan untuk mendapatkan hasil perbandingan ketiga teknik pengukuran adalah data obat pada sebuah apotek di Pekanbaru. Data dikelompokkan berdasarkan kategori, jenis dan satuan obat dengan k uji. Proses pengujian dimulai dari k uji 2 karena pada k uji 1 tidak terlihat hasil anggota cluster karena mempunyai nilai yang sama, dan pengukuran data obat berakhir k 10 karena pada k 10 sudah menghasilkan jumlah

anggota cluster berbeda. Pengujian dilakukan menggunakan RapidMiner. Hasil cluster dioptimalisasi menggunakan DBI dimana k dengan nilai DBI terendah merupakan cluster yang optimal (Davies & Bouldin, 1979) dengan persamaan sebagai berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \quad (4)$$

Langkah akhir adalah melakukan analisis terhadap hasil cluster dari setiap pengukuran numerik.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengolahan *RapidMiner*

Data obat diuji menggunakan aplikasi RapidMiner dengan prinsip kerja yang merujuk pada persamaan (1) hingga persamaan (4). Data obat ditampilkan sebagai berikut:

Tabel 1. Data Obat

Kode Obat	Nama Obat	Kategori	Jenis	Satuan
1	Abocath / Otsucath 24	5	6	1
2	Absolute Fh 60	1	6	2
3	MI New Absolute Ph 150	1	6	2
4	MI New Acarbose Tab 50 Mg	4	1	3
...				
54	Arcamox Dry Syrup 60 Ml	4	5	2
55	As. Mefenamat Kaplet Ogb (Er)	3	1	5
...				
381	Piroxicam 10 Mg 120's Kf/Inf	2	1	3
382	Piroxicam 20 Mg 120's Inf	2	1	3
...				
456	Vitacimin Tablet Sweetlets	1	1	3
...				
462	Woods Herbal 60 Ml	6	5	2
463	Yusimox Syr	5	5	2

Sumber: Apotek X Pekanbaru

Tabel 1 merupakan data obat yang diujikan dengan pengukuran Manhattan, Canberra dan DTWD. Data tersebut telah dikonversi ke dalam bentuk angka sesuai ketentuan yang berlaku pada sistem apotek X

dan akan dikelompokkan berdasarkan Kategori, Jenis dan Satuan.

Selanjutnya data diolah menggunakan RapidMiner yang berakhir di k uji 10, setiap pengukuran numerik menghasilkan *cluster* dengan jumlah anggota *cluster* yang berbeda. Berikut ditampilkan jumlah anggota *cluster* untuk pengukuran *Manhattan Distance*.

Tabel 2. Pengelompokan Manhattan Distance

K Uji	Anggota Cluster
2	cluster_1 (395), cluster_0 (68)
3	cluster_0 (217), cluster_2 (179), cluster_1 (67)
4	cluster_2 (68), cluster_0 (159), cluster_3 (169), cluster_1 (67)
5	cluster_4 (124), cluster_1 (105), cluster_2 (167), cluster_0 (32), cluster_3 (35)
6	cluster_3 (66), cluster_2 (153), cluster_0 (73), cluster_1 (104), cluster_5 (36), cluster_4 (31)
7	cluster_6 (66), cluster_3 (153), cluster_0 (72), cluster_2 (105), cluster_4 (29), cluster_1 (27), cluster_5 (11)
8	cluster_7 (66), cluster_4 (153), cluster_0 (56), cluster_1 (30), cluster_3 (91), cluster_6 (29), cluster_5 (27), cluster_2 (11)
9	cluster_8 (33), cluster_3 (89), cluster_0 (72), cluster_5 (96), cluster_7 (34), cluster_2 (10), cluster_1 (31), cluster_4 (72), cluster_6 (26)
10	cluster_9 (33), cluster_4 (81), cluster_0 (53), cluster_1 (31), cluster_3 (91), cluster_8 (35), cluster_5 (20), cluster_2 (31), cluster_6 (72), cluster_7 (16)
11	cluster_5 (33), cluster_1 (81), cluster_3 (68), cluster_8 (30), cluster_10 (20), cluster_2 (34), cluster_6 (58), cluster_9 (30), cluster_0 (19), cluster_4 (18), cluster_7 (72)

Berdasarkan Tabel 2, terlihat bahwa k uji 2 – 4 memiliki sebaran data yang sama. *Cluster* 0 pada K=2 memiliki kesamaan anggota *cluster* dengan *cluster* 1 pada K=3 dan K=4. Artinya anggota *cluster* tersebut memiliki kesamaan yang erat walaupun k uji berbeda. Semakin banyak jumlah k uji, maka kesamaan yang terkelompokkan semakin spesifik.

Selanjutnya adalah hasil *RapidMiner* yang menampilkan perhitungan DBI setiap pengukuran. Hasil DBI selengkapnya dapat dilihat pada Tabel 3.

Tabel 3. Hasil Perhitungan DBI

K Uji	Manhattan Distance	Canberra Distance	DTWD
2	0,752	2,394	0,868
3	0,756	0,873	1,049
4	0,763	0,934	1,077
5	1,097	1,227	1,063
6	1,635	1,030	1,024
7	0,957	1,574	1,085
8	0,872	2,732	1,347
9	1,194	1,717	1,154
10	1,102	1,708	1,373
11	1,285	2,268	1,093

Pada Tabel 3 memperlihatkan hasil pengelompokkan dengan nilai DBI untuk setiap pengukuran numerik. *Cluster* dengan nilai DBI terendah merupakan *cluster* optimal. Berdasarkan Tabel 3, *cluster* optimal untuk Manhattan dan DTWD terdapat pada k=2 dengan nilai DBI 0,752 dan 0,868. Untuk pengukuran *Canberra*, *cluster* optimal berada pada k=3 dengan nilai DBI sebesar 0,873.

3.2 Analisis Hasil Cluster Optimal

Untuk sebaran data pada setiap *cluster* optimal, *RapidMiner* memperlihatkan sebaran data yang hampir sama. Adapun jumlah anggota *cluster* optimal setiap pengukuran selengkapnya terdapat pada Tabel 4.

Tabel 4. Jumlah Anggota Cluster Optimal

K Optimal	K Optimal	Anggota Cluster
Manhattan	2	cluster_1 (395), cluster_0 (68)
Canberra	3	cluster_1 (226), cluster_0 (64), cluster_2 (173)
DTWD	2	cluster_1 (394), cluster_0 (69)

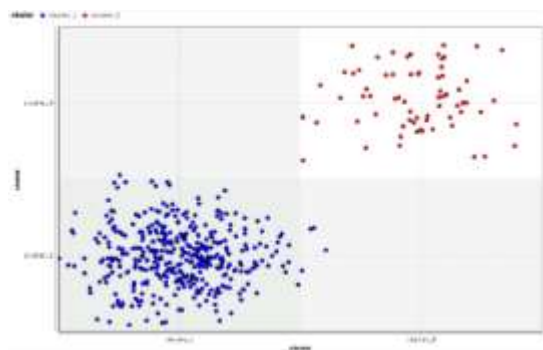
Secara kuantitatif, tidak terlihat perbedaan signifikan terhadap sebaran anggota *cluster* pada *cluster* optimal, seperti yang diperlihatkan Tabel 4. Namun pada kenyataannya terdapat sejumlah data yang tidak sama penempatannya. Berikut diperlihatkan perbandingan data dalam *cluster* optimal pada *Manhattan* dan DTWD.

Tabel 5. Perbandingan Penempatan Anggota Cluster Optimal K=2

Kode Obat	Cluster Manhattan	Kode Obat	Cluster DTWD
42	cluster_0	42	cluster_0
...			
105	cluster_0	105	cluster_1
...			
320	cluster_0	320	cluster_0
321	cluster_0	321	cluster_1
333	cluster_0	333	cluster_0
340	cluster_0	340	cluster_1
348	cluster_0	348	cluster_0
352	cluster_0	352	cluster_1
353	cluster_0	353	cluster_0
...			
398	cluster_0	398	cluster_1
...			
452	cluster_0	452	cluster_0

Berdasarkan Tabel 5, jika dilihat dari cluster 0 untuk k uji 2 pada kedua pengukuran, terdapat 5 data yang berbeda, dimana 5 data tersebut masuk ke dalam cluster 0 pada Manhattan namun berada pada cluster 1 pada pengukuran DTWD. Artinya terdapat 7,35% perbedaan posisi data dalam pengelompokkan. Hal ini terjadi karena perbedaan cara pengukuran. Misalnya *Manhattan* menggunakan nilai *absolut* yang menjadikan perhitungan lebih cepat dan sederhana sehingga kelima data dianggap memiliki kemiripan yang lebih kuat dengan data lainnya dan masuk ke dalam cluster 0.

Sebaran data untuk cluster optimal setiap pengukuran juga ditampilkan *RapidMiner* dalam bentuk grafik pada Gambar 2, 3 dan 4.



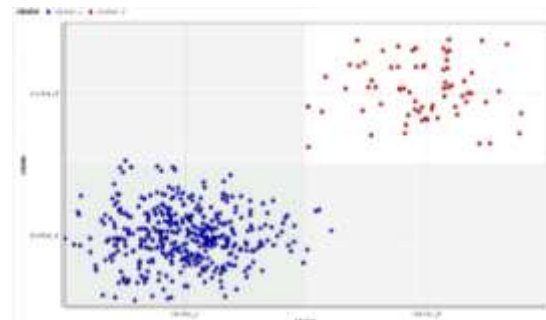
Gambar 2. Sebaran Data Manhattan K=2

Untuk Gambar 2, Grafik memperlihatkan bahwa sebaran anggota cluster 0 lebih dominan. Artinya data memiliki kemiripan yang lebih kuat pada cluster tersebut.



Gambar 3. Sebaran Data Canberra K=3

Pada Gambar 3, grafik memperlihatkan bahwa anggota cluster 1 lebih dominan. Artinya data memiliki kemiripan yang lebih kuat pada cluster tersebut.



Gambar 4. Sebaran Data DTWD K=2

Gambar 2 dan 4 memperlihatkan sebaran data untuk *Manhattan* dan *DTWD* yang memiliki kemiripan seperti yang telah dijelaskan pada Tabel 4 dan 5.

4. SIMPULAN

Hasil dari penelitian ini bertujuan untuk menentukan hasil *clusterisasi* dari pengujian berdasarkan *variabel* yang digunakan untuk data uji seperti kategori, jenis dan satuan obat. Proses *cluster* dilihat dari uji pengukuran *Manhattan*, *Canberra* dan *DTWD* dapat memberikan hasil pengelompokkan yang baik pada data uji dengan cluster optimal berada pada k uji yang masing-masing adalah k=2 di uji dengan *Manhattan*, k=3 diuji dengan *Canberra* dan k=2 diuji dengan *DTWD*. Apabila dilihat dari hasil ketiga pengukuran metode dapat disimpulkan bahwa metode *Manhattan* dengan hasil *Distance* menunjukkan performa lebih baik pada *variabel* kategori, jenis dan satuan obat, karena memiliki nilai DBI terendah, yakni 0,752 yang terlihat pada k uji 2.

5. DAFTAR PUSTAKA

- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Dinov, M., Lorenz, R., Scott, G., Sharp, D. J., Fagerholm, E. D., & Leech, R. (2016). Novel Modeling of Task vs. Rest Brain State Predictability Using a Dynamic Time Warping Spectrum: Comparisons and Contrasts with Other Standard Measures of Brain Dynamics. *Frontiers in Computational Neuroscience, 10*(46), 1–13. <https://doi.org/10.3389/fncom.2016.00046>
- Gautama, W., Purwanto, Y., & Purboyo, T. W. (2015). Analisis Pengaruh Penggunaan Manhattan Distance Pada Algoritma Clustering Isodata (SelfOrganizing Data Analysis Technique) Untuk Sistem Deteksi Anomali Trafik. *E-Proceeding of Engineering, 2*(3), 7404–7411.
- Gie, W., & Jollyta, D. (2020). Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index: Status Covid-19 Wilayah Riau. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS) 2020, 2*(2020), 187–191.
- Irhamni, F., Damayanti, F., Khusnul K, B., & A, M. (2014). Optimalisasi pengelompokan kecamatan berdasarkan indikator pendidikan menggunakan metode clustering dan davies bouldin index. *Seminar Nasional Dan Teknologi UMJ, 11*, 1–6.
- Jendoubi, S., Martin, A., Liétard, L., Yaghlane, B., Jendoubi, S., Martin, A., Liétard, L., Yaghlane, B., Hadji, H., Time, D., Jendoubi, S., Martin, A., & Liétard, L. (2015). Dynamic Time Warping Distance for Message Propagation Classification in Twitter. *HAL, (Juli)*, 419–428.
- Jiang, W., Wang, M., Deng, X., & Gou, L. (2019). Fault diagnosis based on TOPSIS method with Manhattan distance. *Advances in Mechanical Engineering, 11*(3), 1–12. <https://doi.org/10.1177/1687814019833279>
- Jollyta, D., Siddik, M., Mawengkang, H., & Efendi, S. (2021). *Teknik Evaluasi Cluster Solusi Menggunakan Python dan RapidMiner* (Pertama). Deepublish.
- Lance, G. N., & Williams, W. T. (2012). A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems. *The Computer Research Section, 9*(4), 373–380. <https://doi.org/10.1093/comjnl/9.4.373>
- Miftahuddin, Y., Umaroh, S., & Karim, F. R. (2020). Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan. *Jurnal Tekno Insentif, 14*(2), 69–77. <https://doi.org/10.36787/jti.v14i2.270>
- Nishom, M., & Fathoni, M. Y. (2018). Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT), 3*(2), 237–241. <https://doi.org/10.30591/jpit.v3i2.909>
- Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab* (Aldo Saholo (ed.); I). Andi.
- Pulungan, A. F., Zarlis, M., & Suwilo, S. (2019). Analysis of Braycurtis, Canberra and Euclidean Distance in KNN Algorithm. *Sinkron, 4*(1), 74–77. <https://doi.org/10.33395/sinkron.v4i1.10207>
- Putra, I. N. T. A., & Harjoko, A. (2018). Face Recognition Based Mobile Using Fisherface and Distance Classifier. *JST (Jurnal Sains Dan Teknologi), 7*(1), 135–145. <https://doi.org/10.23887/jst-undiksha.v7i1.13267>
- Sitompul, B. J. D., Sitompul, O. S., & Sihombing, P. (2019). Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm. *International Conference on Computing and Applied Informatics 2018, 1235*(1),

1–8. <https://doi.org/10.1088/1742-6596/1235/1/012015>

- Sudarma, M., & Ariyani, S. (2014). Implementation of dynamic time warping method for the vehicle number license recognition. *International Journal of Electrical and Computer Engineering*, 4(2), 278–284. <https://doi.org/10.11591/ijece.v4i2.5169>
- Yang, Y. M., Jia, R., Xun, H., Yang, J., Chen, Q., Zeng, X. G., & Yang, M. (2018). Determining the number of instars in *Simulium quinquestriatum* (Diptera: Simuliidae) using k-means clustering via the canberra distance. *Journal of Medical Entomology*, 55(4), 808–816. <https://doi.org/10.1093/jme/tjy024>
- Zhang, Z., Tavenard, R., Bailly, A., Tang, X., Tang, P., & Corpetti, T. (2017). Length To cite this version: HAL Id: hal-01470554. *HAL Archives-Ouvertes*, 393(February), 91–107.