

KLASTERISASI DESTINASI WISATA BERDASARKAN KOORDINAT GPS DENGAN METODE KERNEL K-MEANS

Yos Heline Laura M Purba ¹⁾ Wasino ²⁾

¹⁾²⁾ Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia
email : yos.825190083@stu.untar.ac.id ¹⁾, wasino@fti.untar.ac.id ²⁾

ABSTRACT

Although *K-Means Clustering*, a sort of non-hierarchical cluster analysis, is frequently employed, it has limitations when processing data with non-linearly separable characteristics (no distinct boundaries) and overlapping clusters, or when visually comparing the results of clusters to those of other clusters. When dealing with data that have overlapping and non-linearly separable cluster features, the Gaussian Kernel function in Kernel *K-Means Clustering* can be applied. The input data must be displayed using kernel functions in a new dimension, which is where Kernel *K-Means Clustering* and *K-Means* diverge.

The purpose of this study is to determine the concept of steps and the results of Kernel *K-Means Clustering* analysis to classify tourist destinations in the *Pesonajawa* database based on GPS coordinates. The data used was obtained from the *PesonaJava.com* database, which consists of 28 tourist destinations located on the island of Java. The silhouette score is used in this study to evaluate the cluster results.

The evaluation of the cluster results based on the research results revealed that the best number of clusters was $K = 6$, with a sigma value of $= 5$.

Key words

RBF kernel, kernel K-Means Clustering, Gps Coordinates, Silhouette Score

1. Pendahuluan

Klasterisasi (*clustering*) merupakan salah satu metode *data mining* yang sering digunakan untuk mengenali banyak data yang akan dibagi menjadi beberapa klaster yang dimana dalam satu klaster data terdapat kemiripan sehingga dimasukkan ke dalam satu kelompok data dan data yang tidak memiliki kemiripan akan dimasukkan ke dalam klaster yang berbeda lagi sehingga dalam satu klaster atau kelompok data hanya memiliki sedikit variasi data. Peran *clustering* dalam ilmu pengetahuan juga semakin meningkat karena perkembangan teknologi yang ada dan meningkatnya kebutuhan manusia yang mau tidak mau harus dihadapi. Kebutuhan yang dimaksud dapat mencakup seluruh

bidang tanpa terkecuali seperti bidang pemerintahan, pendidikan, pemasaran bisnis, dan pariwisata.

Metode analisis *cluster* yang kerap digunakan adalah *K-Means*. Walaupun sering digunakan, metode ini ternyata memiliki kelemahan dalam memproses data yang bersifat non-linear. Data yang bersifat non-linear adalah data yang tidak dapat terpisah secara linear dan tidak dapat dibentuk garis pemisah atau pembatas yang jelas. Cara untuk mengetahui non-linear pada data adalah dengan melihat pada plot hasil akhir *cluster* ada tidaknya objek yang tumpang tindih atau *overlap*. Contoh data *overlap* pada *cluster* adalah saat objek yang dikategorikan sebagai anggota suatu *cluster* terletak diantara *cluster* lainnya, sehingga tidak dapat dibuat garis pembatas linear [1]. Secara signifikan metode kernel *k-means* ini mengurangi waktu komputasi tetapi dengan kerugian kecil dalam kualitas pengelompokan, terutama untuk kumpulan data besar [2]. Ini adalah prosedur iteratif di mana pada awalnya titik data dipetakan dari ruang input ke ruang elemen dimensi yang lebih tinggi melalui transformasi non-linear dan setelah itu meminimalkan kesalahan pengelompokan [3].

Pada pembahasan ini klasterisasi yang akan dibuat ialah klasterisasi destinasi wisata berdasarkan titik koordinat *Global Positioning System* (GPS) dengan menggunakan metode *Kernel K-means*. GPS digunakan sebagai alat untuk menemukan lokasi wisata dan memberikan informasi jarak antar destinasi wisata. Kernel yang digunakan dalam penelitian ini adalah kernel *Gaussian* atau dikenal dengan kernel *Radial Basis Function* (*RBF*) *Gaussian* karena memiliki sifat yang fleksibel dalam penggunaannya [1]. Serta menurut [4], Kernel *RBF Gaussian* adalah kernel yang secara umum dapat digunakan untuk semua jenis data.

Tujuan klasterisasi ini akan mengelompokkan destinasi wisata yang berdekatan sehingga akan terbentuk menjadi satu klaster serta memaksimalkan hasil pembuatan klaster dengan menggunakan metode kernel *k-means clustering*. Data destinasi akan diambil dari database *pesonajawa.com* dimana terdapat beberapa destinasi wisata yang berlokasi di Pulau Jawa.

Metode yang digunakan dalam penelitian ini adalah metode kuantitatif karena dinyatakan dengan angka-angka yang menunjukkan nilai terhadap besaran atau variabel yang diwakilinya. Penelitian kuantitatif

menekankan pada pengujian teori melalui pengukuran variabel penelitian dengan angka dan melakukan analisis data dengan prosedur statistik. Penelitian ini menggunakan pendekatan deduktif yang bertujuan untuk menguji hipotesis. Alat bantu yang digunakan dalam pembuatan aplikasi dan pengolahan data penelitian ini adalah bahasa pemrograman Python, HTML, CSS, JavaScript, dan Flask.

Penelitian yang dilakukan oleh Azizah (2022) yang berjudul “Kernel K-Means Clustering Untuk Pengelompokan Sungai di Kota Semarang Berdasarkan Faktor Pencemaran Air” memiliki kemiripan dengan penelitian ini. Pada penelitian yang dilakukan oleh [1] menghasilkan kesimpulan bahwa nilai variabel yang lebih tinggi dibandingkan dengan jumlah cluster, dapat menyebabkan pencemaran air terburuk jatuh di cluster tersebut.

Penelitian ini bermanfaat untuk meningkatkan pemahaman tentang klusterisasi berdasarkan titik koordinat GPS dengan metode kernel k-means, dapat menjadi masukan dan referensi untuk penelitian yang topiknya berhubungan dengan klusterisasi data.

2. Dasar Teori

2.1 Database

Database (Basis Data) adalah himpunan kelompok data yang saling berhubungan yang diorganisasi sedemikian rupa sehingga kelak dapat dimanfaatkan secara cepat dan mudah. *Database* adalah kumpulan informasi yang disimpan didalam computer secara sistematis sehingga dapat diperiksa dan diakses menggunakan suatu program komputer [5]. *Database* adalah suatu sistem informasi yang mengintegrasikan kumpulan dari data yang saling berhubungan satu dengan lainnya [6].

2.2 Analisis dan Perancangan Sistem Informasi

Analisa dan perancangan sistem informasi (ANSI) merupakan suatu proses penguraian suatu pokok dan menyelidiki keadaan yang sebenarnya dalam sebuah entitas atau guna mencari indikasi komponen dan unsur-unsur penting dalam membangun sebuah sistem informasi [6]. Menurut [7], Analisa dan perancangan sistem informasi adalah upaya untuk mengembangkan pengetahuan, mengembangkan dan menguji teori. Dalam menganalisis dan merancang sebuah system informasi, ada tahapan-tahapan analisis yang dilakukan, Adapun tahapan itu adalah:

1. Memahami kebutuhan dan masalah
2. Menangkap visi
3. Menentukan solusi
4. Membicarakan dan visi beserta solusinya
5. Membangun solusi, atau mengarahkan orang lain dalam membangun solusi

6. Melakukan konfirmasi bahwa solusi yang dibuat sudah sesuai dengan kebutuhan yang ada (melakukan *review*)
7. Meluncurkan aplikasi solusi.

2.3 Machine Learning

Machine learning adalah metode komputasi yang menggunakan pengalaman untuk meningkatkan kinerja atau membuat prediksi yang akurat [8]. *Machine learning* adalah seperangkat teknik yang membantu memproses dan memprediksi data yang sangat besar dengan menyajikan data menggunakan algoritma pembelajaran. Istilah *machine learning* pertama kali didefinisikan oleh Arthur Samuel pada tahun 1959. Menurut Arthur Samuel, machine learning adalah suatu bidang ilmu komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas. Adapun proses pembelajaran yang dimaksud adalah suatu usaha dalam memperoleh kecerdasan yang melalui dua tahap antara lain latihan (*training*) dan pengujian (*testing*) [9].

Dalam pembelajarannya, *machine learning* dibagi menjadi 3 bagian yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*.

1. *Supervised learning* adalah pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.
2. *Unsupervised learning* adalah pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokkan data berdasarkan karakteristik-karakteristik yang ditemui.
3. *Reinforcement learning*, pada tahapan ini pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajar.

2.4 Clustering

Clustering atau klusterisasi adalah metode pengelompokan data. Menurut [1], *clustering* adalah pengelompokan data berdasarkan kesamaan karakteristik antar obyek. *Clustering* merupakan salah satu metode *machine learning* dan termasuk dalam *unsupervised learning*. Tujuan utama dari *clustering* adalah untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengelompokkan objek data, sehingga setiap objek yang memiliki kesamaan terbanyak akan berada dalam *cluster* yang sama. *Cluster-cluster* yang terbentuk dalam satu *cluster* mempunyai ciri yang relatif sama (homogen), sedangkan antar cluster mempunyai ciri yang berbeda (heterogen).

Clustering juga disebut sebagai segmentasi data karena *clustering* berpartisipasi banyak data kedalam banyak kelompok berdasarkan kesamaanya [10].

2.5 Fungsi Kernel

Fungsi kernel adalah fungsi matematika yang digunakan untuk mentransformasikan data ke dalam ruang dimensi yang lebih tinggi, sehingga memungkinkan untuk menggunakan batas keputusan linear untuk mengklasifikasikan data. Distribusi Gaussian memiliki dua parameter, yaitu μ (rata-rata) data, dan σ (standard deviasi atau simpangan baku). Kernel RBF (*Radial Basis Function*) atau sering disebut dengan kernel Gaussian adalah jenis fungsi kernel yang dapat digunakan dengan SVM non-linear. Ini adalah fungsi kernel yang menghitung kemiripan antara dua titik data dengan mempertimbangkan jarak antara mereka dan sering digunakan ketika data tidak dapat dipisahkan secara linear. Menurut [11], kernel digunakan untuk mengatasi permasalahan dimensi.

Metode K-Means memiliki kelemahan dalam pengelompokan data khususnya untuk data yang bersifat *non-linearly separable* [12]. Untuk mengatasi kelemahan tersebut dapat menggunakan fungsi kernel. Kernel bekerja dengan cara memetakan data dari ruang asli ke ruang fitur dimensi yang lebih tinggi. Dengan begitu, objek data yang semula tidak dapat dipisahkan dengan sempurna dapat dipisahkan melalui fitur dimensi yang lebih tinggi tersebut.

Contoh fungsi kernel yang sering digunakan antara lain *Linear*, *Polinomial*, *Radial Basis Function* (RBF), serta *Sigmoid* [10]. Persamaan perhitungan untuk fungsi kernel *Linear*, *Polinomial*, RBF, serta *Sigmoid* masing-masing terdapat pada Persamaan(1), Persamaan(2), Persamaan(3), serta Persamaan(4).

1. Linear

$$k(x_i, x_j) = x_i^T x_j \quad (1)$$

Keterangan:

k = Fungsi kernel

x_i = Data ke $- i$

x_j = Data ke $- j$

2. Polinomial

$$(x_i, x_j) = (x_i^T x_j + r)^d, y > 0 \quad (2)$$

Keterangan:

k = Fungsi kernel

x_i = Data ke $- i$

x_j = Data ke $- j$

r = Konstanta

d = Parameter *degree* (1, 2, 3, ..., n)

3. Radial Basis Function (RBF)

$$k(x_i, x_j) = \exp \left\{ - \frac{\|x_i - x_j\|^2}{2\sigma^2} \right\} \quad (3)$$

Keterangan:

k = Fungsi kernel

x_i = Data ke $- i$

x_j = Data ke $- j$

σ = *Varians*

$$\|x_i - x_j\|^2 = \text{jarak euclidean antara } x_i \text{ dan } x_j$$

4. Sigmoid

$$k(x_i, x_j) = \tanh(x_i^T x_j + r) \quad (4)$$

Keterangan:

k = Fungsi kernel

x_i = Data ke $- i$

x_j = Data ke $- j$

r = Konstanta

γ = Parameter kernel

2.6 Variabel penelitian

Variabel merupakan sesuatu yang menjadi obyek pengamatan penelitian, atau apa yang menjadi perhatian penelitian, yang selanjutnya akan dijadikan obyek didalam menentukan tujuan penelitian [7].

Penelitian ini akan menggunakan koordinat GPS dengan garis lintang dan garis bujur (*Latitude* dan *Longitude*) sebagai variabel dalam melakukan klusterisasi. *Global Positioning System* (GPS) adalah sistem satelit navigasi dan penentuan posisi yang dimiliki dan dikelola oleh Amerika Serikat.

Teknologi GPS digunakan untuk memperkirakan lokasi sebagai nilai bujur dan lintang dalam menemukan lokasi wisata. Destinasi wisata dapat diprediksi secara efisien menggunakan teknologi ini [13].

2.7 Algoritma K-Means

Metode K-Means merupakan metode yang telah banyak diimplementasikan oleh praktisi Teknologi Informasi [10]. Metode K-Means adalah salah satu metode *clustering non-hierarki*. K-Means merupakan bagian dari data mining dengan prinsip *unsupervised learning*. Algoritma K-means adalah algoritma yang terbaik dalam algoritma *partitional clustering* dan yang paling sering digunakan di antara algoritma *clustering* lainnya karena kesederhanaan dan efisiensinya [14].

Dalam melakukan klusterisasi, algoritma K-means mendapatkan sedikit kendala yaitu nilai K sulit diprediksi dan tidak bekerja dengan baik. Ini tidak disarankan untuk digunakan untuk menganalisis data yang banyak dan berukuran besar. Jadi dalam keterbatasan ini, dihasilkanlah metode pengelompokan Kernel k-means [3].

2.8 Algoritma Kernel K-Means

Algoritma Kernel K-Means menerapkan trik yang sama seperti K-Means tetapi dengan satu perbedaan bahwa dalam perhitungan jarak, metode kernel digunakan sebagai pengganti perhitungan jarak Euclidean [10]. Metode ini menggabungkan ide dasar dari k-means dengan fitur yang diberikan oleh kernel. Cara untuk menghitung jarak antar objek pada Kernel K-Means *Clustering* adalah dengan metode jarak Square Euclidean. Jarak Square Euclidean merupakan jarak

yang dikembangkan dari jarak Euclidean. Jarak Square Euclidean jumlah perbedaan deviasi di dalam nilai untuk setiap variabel.

3. Hasil dan Pembahasan

3.1 Clustering Menggunakan Fungsi Kernel

Tahapan pelaksanaan *clustering* Kernel K-Means ialah sebagai berikut [12]:

1. Tentukan jumlah cluster k.
2. Inisialisasi k pusat cluster secara acak
3. Alokasikan semua data ke cluster terdekat berdasarkan jarak yang dihitung menggunakan rumus fungsi pemetaan ϕ dan fungsi kernel.

$$\min \sum_{i=1}^n \sum_{j=1}^m u_{ij} \|\phi(x_i) - \tilde{c}_j\|^2 \quad (5)$$

Dimana \tilde{c}_j merupakan cluster center dan dihitung dengan menggunakan rumus dibawah:

$$\tilde{c}_j = \frac{1}{n_j} \sum_{l=1}^n u_{lj} \phi(x_l) \quad (6)$$

Sehingga:

$$\begin{aligned} \min &= \min \sum_{i=1}^n \sum_{j=1}^m u_{ij} \left\| \phi(x_i) - \frac{1}{n_j} \sum_{l=1}^n u_{lj} \phi(x_l) \right\|^2 \quad (7) \\ &= \min \sum_{i=1}^n \sum_{j=1}^m u_{ij} \left((\phi(x_i) - \frac{1}{n_j} \sum_{l=1}^n u_{lj} \phi(x_l)) (\phi(x_i) - \frac{1}{n_j} \sum_{l=1}^n u_{lj} \phi(x_l)) \right) \\ &= \min \sum_{i=1}^n \sum_{j=1}^m u_{ij} \left((\phi(x_i) \cdot \phi(x_i)) - 2 \left(\frac{\phi(x_i)}{n_j} \sum_{l=1}^n u_{lj} \phi(x_l) \right) + \left(\frac{1}{n_j} \sum_{l=1}^n u_{lj} \phi(x_l) \right)^2 \right) \end{aligned}$$

$$\begin{aligned} &= \min \sum_{i=1}^n \sum_{j=1}^m u_{ij} \left((K(x_i, x_i) \cdot \phi(x_i)) - 2 \left(\frac{1}{n_j} \sum_{l=1}^n u_{lj} K(x_l, x_i) \right) + \left(\frac{1}{(n_j)^2} \sum_{l=1}^n \sum_{p=1}^n (u_{lj})^2 K(x_l, x_p) \right) \right) \end{aligned}$$

Keterangan:

n = Jumlah data

m = Jumlah cluster

K = Fungsi kernel

x_i = Titik data ke-i

x_l = Titik data ke-l

x_p = Titik data ke-p

\tilde{c}_j = *Centroid (cluster center)*

ϕ = Fungsi yang memetakan titik x ke ruang baru yang berdimensi lebih tinggi.

i = Index untuk keseluruhan data

j = Index untuk cluster

l = Index untuk data yang terdapat pada cluster ke-i

p = Index untuk data yang terdapat pada cluster ke-j

u_{ij} = Nilai keanggotaan data ke-i terhadap cluster ke-j; Bernilai 1 apabila merupakan anggota, bernilai 0 jika bukan merupakan anggota dari cluster yang diproses.

u_{lj} = Nilai keanggotaan data ke-l terhadap cluster ke-j

= Fungsi yang memetakan titik x ke ruang baru yang berdimensi lebih tinggi

n_j = Banyaknya data pada cluster ke-j.

4. Setelah mendapat jarak titik data terhadap masing-masing cluster pada langkah 3, jarak terdekat titik data dengan suatu cluster berarti titik data tersebut termasuk dalam cluster tersebut.
5. Perhatikan kondisi berhenti, misal jika pusat cluster tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah.
6. Langkah selanjutnya adalah menentukan apakah diperlukan iterasi selanjutnya atau tidak (*stop condition*). Dalam Langkah ini yang harus dihitung adalah jumlah iterasi maksimum, fungsi obyektif dan *threshold*. Fungsi ini dapat dihitung dengan menggunakan rumus dibawah ini.

$$F = \sum_{j=1}^{n_{data}} \sum_{i=1}^k a_{ij} D(x_j, C_i) \quad (8)$$

$$\text{delta} = |F_{\text{baru}} - F_{\text{lama}}| \tag{9}$$

Keterangan:

- F = Fungsi obyektif
- K = Jumlah cluster
- n_{data} = Jumlah data
- a_{ij} = Nilai keanggotaan data ke-j terhadap cluster ke-i
- $D(x_j, C_i)$ = Jarak antara titik data ke-j terhadap cluster ke-i
- F_{baru} = Fungsi obyektif dari iterasi ke-i
- F_{lama} = Fungsi obyektif dari iterasi ke-(i-1)

Dengan ketentuan sebagai berikut:

Jika $\text{delta} \geq \text{threshold}$, maka iterasi akan berlanjut. Tetapi jika $\text{delta} < \text{threshold}$, maka iterasi berhenti.

3.2 Perhitungan Silhouette Score

Silhouette score bernilai -1 hingga 1. Semakin dekat dengan 1 maka hasil evaluasinya juga akan semakin akurat [10].

Tahapan perhitungan *silhouette score* adalah sebagai berikut:

1. Menghitung rata-rata jarak dari suatu titik i ke titik-titik lainnya yang berada dalam satu *cluster*.
2. Menghitung rata-rata jarak dari titik i ke titik-titik lain yang berada di cluster lainnya, kemudian diambil nilai yang paling kecil
3. *Silhouette score* dapat diketahui melalui perhitungan pada Persamaan

$$S(i) = \frac{b(i) - a(i)}{\text{maks}(a(i), b(i))} \tag{10}$$

Keterangan:

- $a(i)$ = rerata jarak setiap titik data dalam sebuah cluster untuk data ke-i
- $b(i)$ = rerata jarak antar semua cluster untuk data ke-i

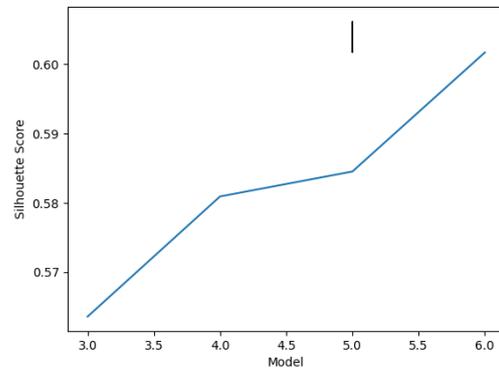
3.3 Penentuan Jumlah Cluster dan Nilai Sigma

Pada penelitian ini, ditentukan banyaknya jumlah *cluster* ada 6 buah dan nilai sigma sama dengan 5. Hal ini didukung dengan pengujian validasi yang sudah dilakukan dengan menggunakan metode *silhouette score*. Tabel 1 menunjukkan hasil perhitungan jumlah *cluster* dan nilai sigma kernel.

Tabel 1 Perbandingan Hasil Perhitungan Silhouette Score

Jumlah Cluster (K)	Nilai Sigma	Nilai Silhouette
3	1	0,563608
4	2	0,580946
5	3	0,584546
6	5	0,601721

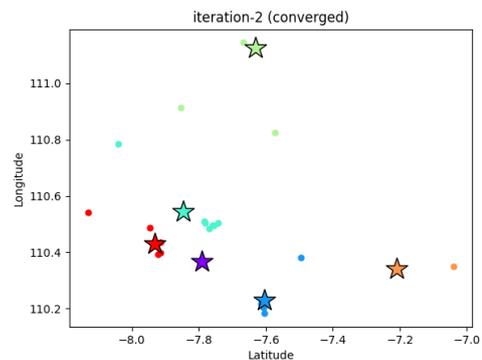
Dari hasil pengujian yang dilakukan dengan jumlah $K = 3, 4, 5,$ dan 6 serta nilai sigma = 1, 2, 3, 5 menghasilkan beragam nilai *silhouette score*. Dari keseluruhan nilai terdapat nilai *silhouette* tertinggi dan mendekati 1 adalah jumlah $k = 6$ dengan nilai sigma = 5 yakni 0,601721. Berikut tampilan perhitungan jumlah dan nilai sigma jika ditampilkan dalam bentuk grafik.



Gambar 1 Grafik Perhitungan Jumlah Cluster dan Nilai Sigma

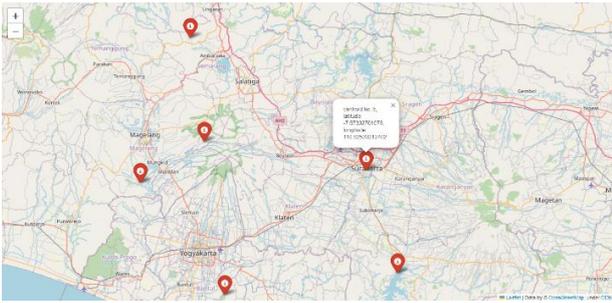
3.4 Hasil Clustering

Setelah proses validasi selesai, berikutnya akan melakukan klasterisasi data koordinat gps dari destinasi wisata. Pada kode klasterisasi ini, jumlah klaster dan nilai sigma harus sudah didefinisikan terlebih dahulu. Serta untuk pemilihan pusat klaster akan dilakukan secara acak. Pada saat proses berjalannya *clustering*, iterasi akan terus bertambah sampai dimana tidak ada lagi *centroid* maupun data klaster yang berpindah. Dibawah ini merupakan plot hasil *clustering* menggunakan *rbf kernel clustering*.

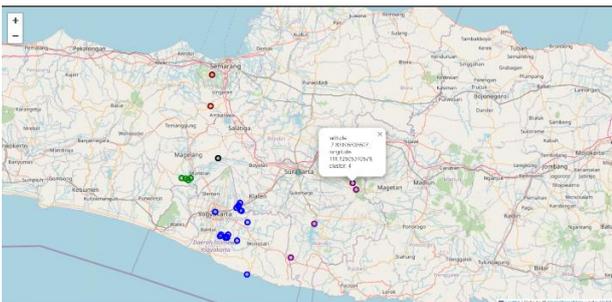


Gambar 2 Hasil Iterasi (Converged)

Setelah proses iterasi sudah *converged*, maka hasil *cluster* dapat dilihat dari *maps* menggunakan *import folium di python*. Dibawah ini merupakan gambar *maps* hasil *clustering*. Pada gambar dibawah terdapat fungsi *pop-up* untuk memberikan informasi mengenai pion data yang ada *map* tersebut.



Gambar 3 Tampilan Centroid Hasil Clustering pada Maps



Gambar 4 Tampilan Data Cluster Hasil Clustering dan Pop-Up

Terdapat 6 jumlah *cluster* dengan nilai sigma sama dengan 5 yang diperoleh dari hasil penelitian ini. Ternyata untuk data yang memiliki jarak yang tidak banyak seperti data yang ada pada penelitian ini, tidak terlalu diperlukan untuk dilakukannya normalisasi data. Data yang harus dilakukan penyederhanaan itu terjadi ketika ada beberapa data yang memiliki jarak yang terbilang cukup jauh dibandingkan dengan data lainnya. Secara sekilas tidak ada perbedaan yang ditampilkan oleh kernel k-means dan k-means *clustering* biasa. Tapi pada saat memproses data mereka tetap memiliki perbedaan.

Berdasarkan pada gambar 3 tersebut, terlihat bahwa anggota kelompok setiap cluster cukup tersebar dan berkumpul membentuk 6 kelompok *cluster*.

4. Kesimpulan

Melalui penelitian dan pengujian yang telah dilakukan dalam klasterisasi destinasi wisata berdasarkan koordinat GPS dengan metode kernel k-means dapat ditarik beberapa kesimpulan sebagai berikut:

1. Jumlah *cluster* terbaik berdasarkan pengukuran dengan metode *silhouette score* ialah 6 *cluster*.
2. Nilai terbaik sigma dalam *rbf kernel* berdasarkan pengukuran dengan metode *silhouette score* adalah 5.
3. Nilai rata-rata terbaik dari Silhouette Score yang diperoleh adalah 0,601721. Nilai Silhouette Score ini cukup mendekati nilai sempurna dalam Silhouette Score disbanding dengan yang lainnya.
4. Jika data yang ingin dikelompokkan tidak memiliki jarak yang terlalu jauh, maka tidak

usah dilakukan penyederhanaan (normalisasi) data.

5. Dengan menggunakan fungsi kernel, maka jumlah iterasi yang terjadi juga akan semakin sedikit sehingga dapat menghemat waktu iterasi.
6. RBF kernel k-means ternyata dapat digunakan untuk mengelompokkan data menurut jarak (*latitude* dan *longitude*).
7. Nilai sigma pada saat melakukan klasterisasi memiliki peran yang penting karena jika nilainya terlalu tinggi atau bahkan terlalu rendah, hasil *cluster* tidak akan bisa maksimal.

Penelitian selanjutnya yang berkaitan dengan metode kernel k-means dalam melakukan *clustering* data masih bisa untuk dikembangkan lagi, seperti menambahi fitur yang ingin di kelompokkan. Variable baru untuk di jadikan sebagai bahan pengujian yang baru seperti suhu, arah mata angin, atau bahkan jika ingin lebih lagi bisa mengembangkan penelitian ini kedalam website atau aplikasi yang memungkinkan untuk membantu banyak orang dalam melakukan klasterisasi wisata.

REFERENSI

- [1] A. N. Azizah, T. Widiharih, A. R. Hakim, D. Statistika, F. Sains, and D. Matematika, “Kernel K-Means Clustering untuk Pengelompokan Sungai di Kota Semarang Berdasarkan Faktor Pencemaran Air,” vol. 11, no. 2, pp. 228–236, 2022, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [2] T. H. Sarma, P. Viswanath, and B. E. Reddy, “Speeding-up the kernel k-means clustering method: A prototype based hybrid approach,” *Pattern Recognit Lett*, vol. 34, no. 5, pp. 564–573, 2013, doi: 10.1016/j.patrec.2012.11.009.
- [3] P. L. Lekshmy and M. Abdul Rahiman, “Hybrid Approach to Speed-Up the Privacy Preserving Kernel K-means Clustering and its Application in Social Distributed Environment,” *Journal of Network and Systems Management*, vol. 28, no. 2, pp. 398–422, Apr. 2020, doi: 10.1007/s10922-019-09511-1.
- [4] R. Indraswari and A. Z. Arifin, “RBF Kernel Optimisation Method with Particle Swarm Optimization on SVM Using The Analysis of Input Data’s Movement,” *Jurnal Ilmu Komputer dan Informasi*, vol. 10, no. 1, p. 36, Feb. 2017, doi: 10.21609/jiki.v10i1.410.
- [5] F. Tri yuniko and F. Kasma Putra, “Penerapan Teknologi Informasi Web Programing untuk Meningkatkan Pelayanan Publik dalam Bidang Kebijakan Administrasi Kependudukan,” 2017.
- [6] N. Azis, “Analisis Perancangan Sistem Informasi,” 2022.
- [7] R. Wijayanti, D. Paramita, M. M. N. Rizal, C. Riza, and B. Sulistyan, *Metode Penelitian Kuantitatif*. 2021.
- [8] M. Mohri, A. Rostamizadeh, and A. Talwalkar, “Foundations of Machine Learning second edition,” 2018.
- [9] G. bin Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: Theory and applications,”

- Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006, doi: 10.1016/j.neucom.2005.12.126.
- [10] E. F. Achmal, I. Cholissodin, and P. P. Adikara, “Segmentasi Pelanggan menggunakan Metode Kernel K-Means (Studi Kasus: Smartlegal.id),” 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] K. Aprianto, B. Pusat, S. Provinsi, S. Barat, J. Re, and M. N. 10, “Optimasi Kernel K-Means dalam Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia di Indonesia,” 2018.
- [12] E. Riyandani, “‘Big Data vs Big Information vs Big Knowledge’ Oleh: Imam Cholissodin,” 2016. [Online]. Available: <http://bit.ly/2x8ta9S>
- [13] K. Ravikumar and A. RajivKannan, “An enhancement of location estimation and disaster event prediction using density based SPATIO-temporal clustering with GPS,” *Multimed Tools Appl*, vol. 79, no. 5–6, pp. 3929–3941, Feb. 2020, doi: 10.1007/s11042-019-7583-7.
- [14] A. Almayda and S. Saepudin, “Penerapan Data Mining K-Means Clustering untuk Mengelompokkan Berbagai Jenis Smartphone,” 2021.

Yos Heline Laura M. Purba, mahasiswa tingkat akhir Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jakarta.

Wasino S.Kom., M.Kom., Dosen Program Studi Sistem Infomasi, Fakultas Teknologi Informasi, Universitas Tarumangara, Jakarta.