



INNOVATIVE: Journal Of Social Science Research

Volume 4 Nomor 1 Tahun 2024 Page 8846-8854

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Data Gangguan PT. PLN UID Kalselteng

Tri Amri Wijaya^{1✉}, Ema Utami², Hanif Al Fatta³

Universitas Amikom Yogyakarta

Email: triamri@gmail.com^{1✉}

Abstrak

Perusahaan Listrik Negara (PLN) merupakan salah satu perusahaan Badan Usaha Milik Negara, Dalam pendistribusian listrik sering terjadi gangguan yang menyebabkan pasoka listrik sering terdampak diantaranya terjadi pemadaman, Bentuk pencegahan gangguan dilakukan dengan cara pengelompokan titik-titik gangguan yang sering terjadi, Untuk mengatasi permasalahan Terdapat berbagai metode yang dapat digunakan dalam pengelompokan data, diantaranya : DBSCAN dan K-Means. Tujuan dari penelitian ini untuk melakukan pengelompokan data gangguan menggunakan metode DBSCAN dan K-Means. Percobaan dilakukan menggunakan dataset sebanyak 2000 data yang terdiri dari 2 atribut: latitude dan longitude. Berdasarkan penelitian ini, Setelah melalui beberapa percobaan diperoleh hasil bahwa K-Means lebih unggul daripada DBSCAN dalam mengelompokkan data gangguan. Algoritma K-Means memiliki nilai SI terbaik sebesar 0,581 yang terletak pada percobaan dengan nilai $k = 6$.

Kata Kunci: *PLN, Clustering, DBSCAN, K-Means, Silhouette Index*

Abstract

The State Electricity Company (PLN) is one of the state-owned companies. In the distribution of electricity, disturbances often occur which cause the electricity supply to be frequently affected, including blackouts. The form of preventing disturbances is done by grouping disturbance points that frequently occur. To overcome the problems there are various methods that can be used in grouping data, including: DBSCAN and K-Means. The aim of this research is to group disturbance data using the DBSCAN and K-Means methods. The experiment was carried out using a dataset of 2000 data consisting of 2 attributes: latitude and longitude. Based on this research, after several experiments the results were obtained that K-Means was superior to DBSCAN in classifying disturbance data. The K-Means algorithm has the best SI value of 0.581 which is located in experiments with a value of $k = 6$.

Keywords: *PLN, Clustering, DBSCAN, K-Means, Silhouette Index*

PENDAHULUAN

PLN (Perusahaan Listrik Negara) merupakan salah satu perusahaan BUMN (badan usaha milik negara). Dalam pendistribusian listrik sering terjadi gangguan yang menyebabkan pasokan listrik sering terdampak diantaranya terjadi pemadaman. Gangguan yang terjadi dikarenakan faktor internal dan eksternal yang menyebabkan berkurangnya pasokan. Salah satu bentuk pencegahan gangguan dengan melakukan pemantauan terhadap titik-titik gangguan yang dapat digunakan sebagai indikasi terjadinya gangguan,

Dengan karakteristik gangguan pengelompokan data gangguan merupakan salah satu solusi untuk mengetahui lokasi gangguan yang sering terdampak. Terdapat berbagai metode yang dapat digunakan dalam pengelompokan data, diantaranya: DBSCAN dan K-Means. Belum diketahuinya kinerja kedua metode ini, sehingga perlu pemahaman yang baik ketika akan menggunakannya.

Clustering adalah suatu teknik dalam data mining yang bertujuan untuk mengelompokkan data- data (objek) ke dalam beberapa cluster atau kelompok sehingga objek yang serupa disatukan kedalam cluster yang sama, dan yang berbeda harus menjadi bagian dari cluster yang berbeda. Berdasarkan perspektif optimasi, tujuan utama dari clustering untuk memaksimalkan homogenitas (internal) dalam sebuah cluster dan heterogenitas (eksternal) di antara cluster yang berbeda.

Pada penelitian Kamila, Khairunnisa, dan Mustakim (2019) menerapkan metode clustering mengenai perbandingan algoritma K-Means dan K-Medoids untuk pengelompokan data transaksi bongkar muat di Provinsi Riau yang pada kesimpulannya tidak menunjukkan adanya perbedaan yang signifikan terkait pengelompokan data. Kemudian penelitian oleh Mustakim dkk. (2021) kombinasi algoritma DBSCAN dan K-

Medoids untuk mengklasterisasi opini public mengenai bencana alam di Indonesia. Pada penelitian ini algoritma DBSCAN memiliki validitas cluster dengan nilai Silhouette Index (SI) tertinggi yaitu 0,9140 dalam waktu eksekusi rata-rata 83,40 detik. Sementara algoritma K-Medoids memiliki nilai SI hanya 0,2258 dengan waktu eksekusi rata-rata 849,93 detik. Hasil dari penelitian ini frekuensi gempa mendominasi kategori positif, kata bencana mendominasi negatif, dan kata banjir mendominasi kategori netral.

DBSCAN mampu menciptakan beberapa cluster yang berbentuk bebas serta acak (tidak bulat) dan dapat menciptakan cluster dengan lebih mudah jika terdapat noise terhadap beberapa cluster tersebut. Algoritma DBSCAN dapat menemukan setiap cluster dengan bentuk apapun serta efektif mengidentifikasi titik-titik noise yang ada. Sedangkan K-Means merupakan salah satu algoritma yang mudah digunakan dalam untuk melakukan klasterisasi data dalam jumlah besar dan data outlier dengan sangat cepat berdasarkan tingkat kemiripan dan ketidakmiripan objek.

Berdasarkan uraian di atas dan di dukung beberapa penelitian sebelumnya, maka pada penelitian ini akan dilakukan perbandingan algoritma DBSCAN dan K-Means untuk klasterisasi data gangguan dengan mempertimbangkan nilai validitas cluster terbaik sebagai penentu keakuratan algoritma yang digunakan.

METODE PENELITIAN

Pada penelitian ini akan digunakan proses pengambilan data, pre-processing data, dan metode clustering yang meliputi K-Means dan DBSCAN.

Pengambilan Data

Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari PT. PLN UID Kalselteng. Data yang digunakan merupakan data titik gangguan yang terdapat pada wilayah PT. PLN UID Kalselteng pada tahun 2020 – 2022. Data tersebut berisikan 2000 data dan 2 variabel latitude dan longitude sebagai titik gangguan.

Pre-Processing Data

Dalam analisis ini, terdapat dua tahapan pre-processing data yang digunakan. Pre-processing data tersebut meliputi cleaning data dan standarisasi data

- a. Cleaning data dilakukan dengan melengkapi data dan menghapus data kosong ataupun data yang tidak digunakan.
- b. Standarisasi data dilakukan untuk menyamakan data yang tidak sesuai dan untuk menyamakan variabel agar memiliki distribusi yang serupa.

Clustering

1. DBSCAN. Density-Based Spatial Clustering Algorithm with Noise (DBSCAN) adalah algoritma untuk clustering data berdasarkan kepadatan. Konsep kepadatan yang dimaksud dalam DBSCAN adalah jumlah minimal data dalam radius Eps, data tersebut termasuk dalam kategori kepadatan yang diinginkan.

Terdapat dua parameter penting untuk clustering dengan menggunakan algoritma DBSCAN, yakni:

- a. Epsilon (Eps) merupakan jarak maksimal antara dua titik yang diizinkan dalam satu cluster.
- b. Minimum Points (MinPts) adalah jumlah minimal titik sehingga dapat dikatakan sebuah cluster.

Terdapat tiga jenis tipe titik dalam proses clustering dengan menggunakan algoritma DBSCAN, yakni:

- a. Core point adalah titik yang berada dalam satu cluster dan memiliki titik tetangga sedikitnya sebanyak MinPts.
 - b. Border point adalah titik yang berada dalam suatu cluster namun jumlah titik tetangganya kurang dari MinPts.
 - c. Noise point adalah titik yang tidak termasuk anggota cluster manapun.
2. K-Means. K-Means merupakan algoritma clustering yang mengarah pada pembagian n objek pengamatan ke dalam k kelompok (cluster) dan setiap objek pengamatan dimiliki oleh suatu kelompok dengan rata-rata (mean) terdekat. Prinsip utama dari teknik ini adalah menyusun k buah mean dari sekumpulan data. Pada algoritma K-Means, dibutuhkan jarak setiap data input terhadap masing-masing centroid dengan menggunakan jarak Euclid. Jarak Euclid merupakan jarak yang umumnya digunakan pada clustering. Menurut Ediyanti, dkk. jarak Euclid didefinisikan dengan

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, \dots, n \quad (1)$$

dimana

(x, y) : jarak Euclid data x dan y

x_i : objek x ke i

y_i : data y ke i

n : banyaknya objek

Jarak terpendek antara centroid dengan objek menentukan titik cluster antar-objek. Adapun centroid didefinisikan dengan

$$v = \frac{\sum_{i=1}^n x_i}{n}, i = 1, 2, \dots, n \quad (2)$$

Dimana

v : centroid data

x_i : objek x ke i

n : banyaknya objek anggota cluster.

HASIL DAN PEMBAHASAN

Hasil penelitian membahas mengenai proses analisis dengan menggunakan algoritma DBSCAN dan K-Means. Lalu membandingkan antara kedua metode tersebut mana yang lebih efisien secara keseluruhan dengan beberapa perhitungan dengan parameter masing-masing di setiap metodenya.

Pengumpulan dan Preprocessing Data

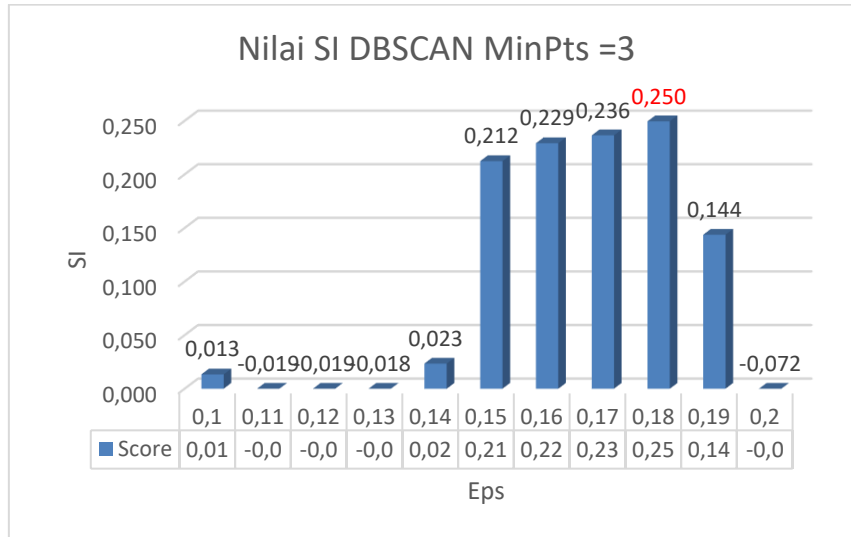
Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari PT. PLN UID Kalselteng. Data yang diambil merupakan data titik gangguan pada wilayah Kalimantan Selatan dan Kalimantan Tengah dari tahun 2020 sampai dengan 2022 yang terdapat 2000 titik gangguan yang dimana titik gangguan tersebut digunakan sebagai parameter perbandingan pengelompokan pada dua metode yang di gunakan yaitu algoritma DBSCAN dan K-Means. Setelah data sudah terkumpul, langkah selanjutnya adalah melakukan preprocessing data. Dalam tahap preprocessing data ini akan dilakukan data cleaning, seleksi data, dan standarisasi data yang dimana data-data yang tidak diinginkan akan dihilangkan sehingga akan menampilkan data yang valid dengan variable latitude dan longitude. Sehingga akan tampil data seperti Tabel 1.

Tabel 1. Data Titik Gangguan hasil cleaning

	lat	lng
0	-2.109758	115.172821
1	-3.389100	114.784230
2	-1.633120	115.044330
3	-2.379460	115.543500
4	-2.948710	115.300430

Clustering dengan Algoritma DBSCAN

Pada Proses klusterisasi menggunakan algoritma DBSCAN menggunakan parameter percobaan cluster sebanyak 10 kali dimana setiap percobaan menggunakan nilai epsilon (Eps) dan MinPoints (MinPts) yang berbeda. Nilai Eps yang digunakan pada penelitian ini yaitu pada rentang 0,1 hingga 0,2, sementara nilai MinPts 3, selanjutnya dilakukan uji menggunakan parameter tersebut, setelah itu untuk validitas cluster menggunakan Silhouette Index (SI). Nilai SI ini di dapat berdasarkan hasil klusterisasi dari hasil hitung menggunakan DBSCAN dapat dilihat pada Gambar 1.

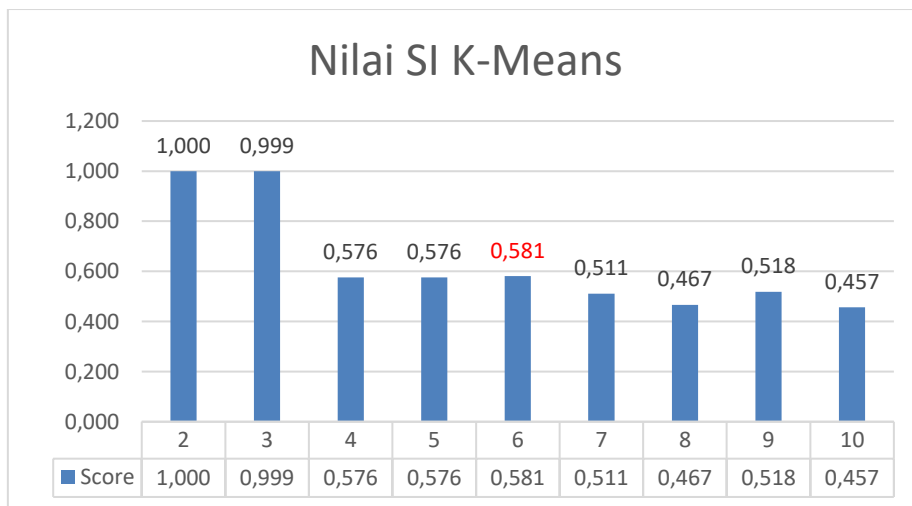


Gambar 1. Grafik Nilai SI hasil klusterisasi DBSCAN

Dapat dilihat hasil nilai SI bahwa nilai SI terbaik pada hasil dari klusterisasi tersebut, pada algoritma DBSCAN terletak pada Eps 0,18. Dengan nilai SI yang diperoleh sebesar 0,250 dari nilai Eps yang lain sehingga nilai tersebut dinyatakan sebagai nilai paling baik diantara Eps lainnya.

Klusterisasi dengan Algoritma K-Means

Pada proses klusterisasi selanjutnya menggunakan algoritma K-Means dengan melakukan percobaan cluster 2 sampai dengan 10, Setelah nilai klusterisasi didapat kemudian dilakukan validitas menggunakan Silhouette Index (SI). Nilai SI ini di dapat berdasarkan hasil klusterisasi dari hasil hitung berdasarkan hasil klusterisasi data menggunakan algoritma K-Means dapat dilihat Gambar 2 berikut.

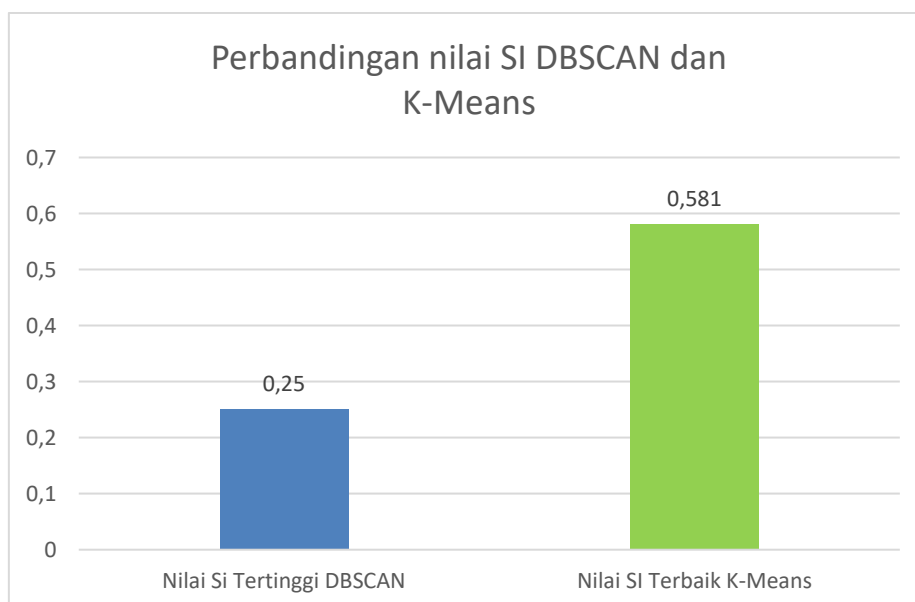


Gambar 2. Grafik nilai SI hasil klusterisasi K-Means

Dapat dilihat hasil nilai SI bahwa nilai SI terbaik pada hasil dari klusterisasi tersebut, pada algoritma K-Means terletak pada cluster 6. Dengan nilai SI diperoleh sebesar 0,581 dari nilai cluster lainnya.

Perbandingan Algoritma DBSCAN dan K-Means

Setelah kita peroleh hasil perhitungan setiap algoritma DBSCAN dan K-Means, maka akan kita bandingkan nilai terbaik pada 2 metode tersebut perbandingan dilihat dari nilai Silhouette Index (SI) yang di peroleh pada 2 metode tersebut. Lalu di bandingkan antara 2 nilai SI terbaik dalam bentuk Grafik nilai SI tertinggi pada setiap algoritma dapat dilihat pada Gambar 3 berikut.



Gambar 3. Grafik perbandingan nilai SI DBSCAN dan K-Means

Hasil Pengujian cluster pada algoritma DBSCAN dan K-Means menggunakan nilai Silhouette Index (SI). klusterisasi pada algoritma DBSCAN yang memiliki nilai SI sebesar

0.250 dengan nilai Eps 0,18. Sementara itu hasil dari klasterisasi menggunakan algoritma K-Means, nilai SI terbaik diperoleh pada percobaan k=6 dengan nilai sebesar 0,581. Maka jika di bandingkan kedua metode tersebut pada penelitian ini, algoritma K-Means memiliki nilai validitas cluster yang lebih baik dibandingkan dengan algoritma DBSCAN dengan selisih yang cukup signifikan yaitu sekitar 0,331.

SIMPULAN

Berdasarkan hasil penelitian dan pengolahan data yang telah dilakukan pada 2 algoritma DBSCAN dan K-Means, pengolahan data menggunakan algoritma DBSCAN diperoleh cluster terbaik pada parameter Eps 0,18 dan MinPts 3 dengan nilai Silhouette Index (SI) sebesar 0.250, sementara itu pada algoritma K-Means diperoleh pada percobaan cluster k = 6 dengan nilai Silhouette Index sebesar 0,581. Sehingga hasil dari klasterisasi tersebut sudah tervalidasi menggunakan Silhouette Index (SI). Berdasarkan hasil pengujian tersebut pada algoritma DBSCAN dan K-Means, maka dapat di simpulkan bahwa hasil dari penelitian ini K-Means lebih efisien dari pada DBSCAN dengan nilai SI terbaik yaitu 0,581 dengan nilai k = 6.

DAFTAR PUSTAKA

- A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia Comput. Sci.*, vol. 148, pp. 291– 302, 2019, doi: 10.1016/j.procs.2019.01.022.
- I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 012016, 2021, doi: 10.1088/1742-6596/1783/1/012016.
- W. Jing, C. Zhao, and C. Jiang, "An improvement method of DBSCAN algorithm on cloud computing," *Procedia Comput. Sci.*, vol. 147, pp. 596–604, 2019, doi: 10.1016/j.procs.2019.01.208.
- B. M. Metisen and H. L. Sari, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110-118. ISSN: 1858 –2680, 2015.

- B. S. Ashari, S. C. Otniel, and R. Rianto, "Perbandingan Kinerja K-Means dengan DBSCAN untuk Metode Clustering data Penjualan Online Retail," *J. Siliwangi Seri Sains dan Teknol.*, vol. 5, no. 2, pp. 64–67, 2019.
- Ediyanti, M. N. Mara, and N. Satyahadewi, Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis, *Buletin Ilmiah Mat. Stat. Dan Terapannya*, vol. 2 no. 2, p 133-136. 2013.