



PENERAPAN ALGORITMA K-MEANS, DBSCAN, DAN AHC UNTUK CLUSTERING KUALITAS GARAM PADA PT. GARAM (PERSERO)

Sigit Susanto Putro¹⁾, Mohammad Syarief²⁾, Eka Mala Sari Rochman³⁾

^{1), 2), 3)} Teknik Informatika Universitas Trunojoyo
Jl Raya Telang Po Bax 2 kamal
Email : sigit.putro@trunojoyo.ac.id

Dikimkan: 17 Desember 2023

Direvisi: 21 Agustus 2024

Diterima: 28 Oktober 2024

Abstrak

Garam merupakan barang hasil produksi rakyat yang berperan penting dalam memenuhi kebutuhan konsumen dan berbagai kegiatan industri. Kualitas garam dapat mempengaruhi berbagai aspek, termasuk kesehatan, cita rasa makanan, dan penggunaan dalam kegiatan industri. Kualitas garam yang buruk dapat mempengaruhi kualitas produk akhir yang dihasilkan. Oleh sebab itu, perlu dilakukan pengelompokan kualitas garam untuk memastikan bahwa garam yang digunakan sesuai dengan kebutuhan dan standar kualitas untuk kebutuhan tertentu. Berkaitan dengan tujuan tersebut penelitian ini menerapkan 3 metode berbeda yaitu K-means, DBSCAN, dan AHC. K-means adalah algoritma clustering yang membagi data ke dalam K kelompok dengan cara meminimalkan jarak antara titik data dan pusat cluster. Agglomerative Hierarchical Clustering adalah metode dalam analisis data yang mengelompokkan objek-objek berdasarkan kesamaan karakteristik dengan cara menggabungkan kelompok-kelompok secara hirarki. DBSCAN adalah algoritma clustering yang menggunakan kerapatan spasial untuk mengelompokkan data. Dari jumlah data sebanyak 350 dengan 9 fitur yang berasal dari PT. Garam Sumenep yang di kelompokkan menggunakan tiga metode dilakukan pengujian kualitas clustering menggunakan silhouette coefficient yang menghasilkan nilai 0.345 untuk metode K-means, 0.32 untuk metode AHC dan 0.5 untuk metode DBSCAN. Hal yang perlu dilakukan supaya hasil cluster dapat maksimal adalah dengan proses penggalan variasi variabel data yang berbeda.

Kata kunci: Garam, Cluster, K-means, DBSCAN, AHC, Madura.

Abstract

Salt is a product produced by the people which plays an important role in meeting consumer needs and various industrial activities. The quality of salt can affect various aspects, including health, food taste, and use in industrial activities. Poor salt quality can affect the quality of the final product produced. Therefore, it is necessary to group the quality of salt to ensure that the salt used meets the needs and quality standards for certain needs. In connection with the aim of this research, 3 different methods are applied, namely K-means, DBSCAN, and AHC. K-means is a clustering algorithm that divides data into K groups by minimizing the distance between data points and the cluster center. Agglomerative Hierarchical Clustering is a method of data analysis that groups objects based on similar characteristics by combining groups hierarchically. DBSCAN is a clustering algorithm that uses spatial density to group data. From a total of 350 data with 9 features originating from PT. Sumenep salt was grouped using three methods which carried out clustering quality measurements using the silhouette coefficient which produced a value of 0.345 for the K-means method, 0.32 for the AHC method and 0.5

for the DBSCAN method. What needs to be done so that cluster results can be maximized is the process of extracting variations in different data variables.

Keyword: Salt, Cluster, K-means, DBSCAN, AHC, Madura.

PENDAHULUAN

Pulau Madura, yang terletak di sebelah timur Pulau Jawa dan dipisahkan oleh Selat Madura, memiliki keadaan geografis yang mendukung perannya sebagai pulau maritim di Indonesia. Lahan pesisir yang luas dan datar memungkinkan Pulau Madura untuk menjadi salah satu pulau penghasil garam terbesar di Indonesia [1]. Garam merupakan barang hasil produksi rakyat yang berperan penting dalam memenuhi kebutuhan konsumsi masyarakat dan berbagai kegiatan industri. Penggunaan garam di berbagai kegiatan industri termasuk, pengolahan makanan, kimia, farmasi, dan lain sebagainya. Kualitas garam yang buruk dapat mempengaruhi kualitas produk akhir yang dihasilkan [2].

PT Garam Persero tersebar di beberapa wilayah pulau madura, salah satunya di Kabupaten Sumenep. PT Garam banyak memproduksi berbagai jenis garam untuk memenuhi berbagai keperluan akan garam baik untuk kebutuhan rumah tangga, industri, peternakan dan pertanian [3]. namun industri garam di indonesia masih menemui beberapa masalah dan hambatan. kualitas garam yang belum maksimal merupakan satu dari sekian banyak masalah pada industri garam [3].

Clustering adalah teknik analisis data yang digunakan untuk mengelompokkan objek-objek data yang memiliki kesamaan dalam suatu set atribut atau karakteristik tertentu ke dalam kelompok-kelompok atau cluster[4]. Tujuan dari clustering adalah untuk mengidentifikasi pola atau struktur yang ada dalam data yang mungkin sulit terlihat secara langsung. Dengan membagi data menjadi kelompok-kelompok yang homogen, clustering membantu dalam memahami karakteristik dan hubungan antar objek dalam dataset yang kompleks, memfasilitasi pengambilan keputusan yang lebih baik, serta membantu dalam eksplorasi data dan pengenalan pola baru [4], [5].

K-means juga memiliki kekurangan, yaitu tidak ada ketentuan yang pasti kapan menentukan pusat awal terbaik dari cluster, jika penentuan pusat awal cluster yang berbeda akan menghasilkan keanggotaan yang berbeda [6]. Pada tahun 2022 dilakukan penelitian yang serupa menggunakan data TB dimana peneliti menggunakan 2 metode berbeda yaitu K-means dan juga FCM. Setelah diuji menggunakan Silhouette Coefficient, K-means mendapatkan hasil sebesar 0.6318 dimana nilai tersebut lebih baik daripada metode FCM [7].

Pada penelitian lain, dilakukan penelitian mengenai kejahatan berdasarkan provinsi yang sering terjadi kejahatan, seperti pencurian, perampokan, penipuan, penganiayaan fisik, dan pembunuhan. Pengelompokan daerah rawan terjadi kejahatan dianggap penting dilakukan untuk bisa dilakukan pencegahan seperti meningkatkan keamanan suatu daerah. Pada penelitian tersebut digunakan dua metode pengelompokan yaitu K-means dan AHC dengan menggunakan complete linkage. Data yang digunakan dari 34 provinsi dengan rentang tahun 2019-2021 dan didapatkan 2 cluster pada K-means masing-masing 17 dan 2 cluster juga pada AHC tetapi ada beberapa anggota cluster yang berbeda [8].

Berdasarkan penelitian sebelumnya yang telah dilakukan dengan menggunakan teknik Clustering, tujuan penelitian ini adalah untuk mengelompokkan kualitas garam pada PT Garam (Persero) dengan beberapa metode cluster seperti K-means, AHC dan DBSCAN. Dimaksudkan agar bisa mengetahui hasil cluster terbaik yang terbentuk dari beberapa metode clustering tersebut.

METODE PENELITIAN

Data

Ada beberapa fitur atau atribut yang digunakan untuk menentukan kualitas garam, diantaranya:

1. Kadar Air: Kadar air dalam garam mengacu pada persentase air yang terkandung di dalam garam. Kadar air yang rendah cenderung diinginkan dalam garam berkualitas baik, karena kelembaban yang tinggi dapat mempengaruhi stabilitas dan umur simpan garam.
2. Tak Larut: Fitur ini mungkin mengukur jumlah komponen yang tidak larut dalam garam. Kadar tak larut yang rendah dapat menunjukkan kualitas garam yang lebih baik, karena garam yang lebih bersih memiliki sedikit kontaminan atau material yang tidak larut.
3. Kalsium dan Magnesium: Kalsium dan magnesium adalah mineral yang dapat mempengaruhi rasa dan kualitas garam. Kadar yang terlalu tinggi dari kedua mineral ini dalam garam dapat memberikan rasa yang kurang diinginkan.
4. Sulfat: Sulfat adalah ion yang juga dapat mempengaruhi rasa dan kualitas garam. Sulfat yang berlebihan dalam garam dapat memberikan rasa yang pahit.
5. NaCl (wb) dan NaCl (db): Persentase garam sodium klorida (NaCl) dalam bentuk basah (wb) dan kering (db) menggambarkan jumlah garam asli dalam sampel garam. NaCl adalah komponen utama garam dan persentase yang lebih tinggi biasanya mengindikasikan garam berkualitas baik.

Preprocessing Data

Preprocessing data bertujuan untuk meningkatkan kualitas dan efektivitas analisis yang dijalankan. Dengan melaksanakan langkah-langkah preprocessing data, kami bertujuan untuk mengatasi berbagai isu yang mungkin timbul dari data mentah, seperti kehadiran nilai yang hilang atau missing value, atau perbedaan skala antara fitur-fitur.

Fitur selection adalah salah satu langkah penting dalam preprocessing data yang bertujuan untuk memilih subset fitur yang paling relevan dan berkontribusi terhadap tujuan analisis data. Tujuan dari fitur selection adalah untuk mengurangi dimensi data, menghilangkan fitur yang tidak penting, serta meningkatkan kualitas dan efisiensi analisis. Dengan mengurangi jumlah fitur, kita dapat menghindari masalah curse of dimensionality dan meningkatkan performansi algoritma analisis data.

Min-Max Normalization adalah teknik yang digunakan untuk mentransformasikan fitur-fitur data ke dalam rentang nilai tertentu, biasanya antara 0 dan 1. Tujuan dari normalisasi ini adalah untuk mengubah skala fitur sehingga setiap fitur memiliki kontribusi yang seimbang dalam analisis, terlepas dari rentang nilai awal yang berbeda-beda. Hal ini juga membantu menghindari dominasi fitur dengan rentang nilai yang besar terhadap fitur lainnya [9]. Dalam Min-Max Normalization, setiap nilai fitur diubah menjadi nilai baru sesuai dengan rumus berikut [10]:

$$norm = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Keterangan :

X : nilai asli fitur

min (x) : nilai minimum dari semua nilai fitur dalam dataset

max (x): nilai maksimum dari semua nilai fitur dalam dataset

Clustering

Clustering merupakan teknik analisis data yang bertujuan untuk mengelompokkan objek-objek data menjadi kelompok-kelompok atau cluster berdasarkan kemiripan karakteristik atau atribut tertentu. Tujuan utama dari clustering adalah untuk mengidentifikasi pola atau struktur tersembunyi dalam data yang mungkin sulit terlihat secara langsung [5]. Dalam proses clustering, objek-objek data yang memiliki kemiripan dalam atribut yang dianalisis dikelompokkan bersama dalam cluster yang sama, sementara objek-objek yang memiliki perbedaan signifikan ditempatkan dalam kelompok yang berbeda. Teknik clustering umumnya menggunakan metode matematis, statistik, atau algoritma komputasi yang berupaya meminimalkan perbedaan antara objek dalam suatu cluster dan memaksimalkan perbedaan antara cluster yang berbeda. Hasil dari clustering dapat membantu mengungkap pola-pola, segmentasi data, serta memberikan wawasan yang bermanfaat dalam berbagai bidang seperti analisis bisnis, pengelompokan objek, atau pemahaman karakteristik data yang kompleks [5].

Setelah kita mengelompokkan data menjadi cluster, langkah selanjutnya adalah menentukan seberapa mirip atau berbedanya objek-objek dalam cluster tersebut. Euclidean distance menjadi alat yang berguna dalam mengukur jarak antara objek-objek tersebut dalam ruang dimensi yang relevan. Prosesnya melibatkan menghitung jarak Euclidean antara setiap pasangan objek dalam cluster dengan mengukur perbedaan atau selisih pada setiap atribut atau fitur yang dianalisis. Semakin kecil nilai Euclidean distance antara dua objek, semakin mirip atau serupa mereka dalam hal atribut yang diamati. Dengan menggunakan Euclidean distance, kita dapat mengukur tingkat kesamaan antara objek-objek dalam cluster dan mengidentifikasi apakah objek tersebut saling mendekati atau jauh berdasarkan karakteristik yang dinilai [5].

K-means

K-means adalah metode clustering yang digunakan untuk mengelompokkan data menjadi beberapa kelompok berdasarkan kesamaan fitur atau atribut [11]. Tujuannya adalah untuk mengidentifikasi pola-pola tersembunyi dalam data dengan membagi objek-objek ke dalam kelompok yang mirip [12]. Langkah-langkah dasar K-means melibatkan inisialisasi pusat-pusat cluster secara acak atau berdasarkan pemilihan awal, kemudian melakukan iterasi untuk menghitung jarak Euclidean antara setiap objek dan pusat cluster terdekat. Objek akan dikelompokkan berdasarkan pusat cluster terdekat, dan titik pusat akan diperbarui dengan mengambil rata-rata atribut dari semua objek dalam kelompok. Proses ini diulang hingga pusat cluster tidak lagi berubah secara signifikan atau mencapai batas iterasi yang ditentukan [11].

Dalam perhitungan jarak Euclidean dalam K-means, yang mengukur seberapa dekat atau jauh dua objek dalam ruang fitur, rumus yang digunakan adalah [11]:

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

Keterangan:

$d(x, y)$ = Nilai jarak antara objek data x dan y .

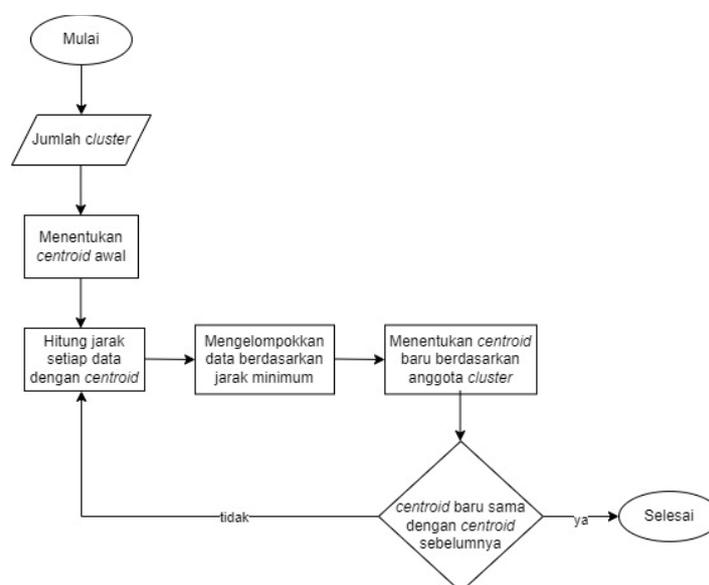
x_i = Data x ke i

y_i = Data y ke i

Metode *clustering K-means* adalah pendekatan yang efektif untuk mengelompokkan data ke dalam beberapa *cluster* yang serupa. Langkah-langkah proses pengelompokan menggunakan algoritma *K-means* dijelaskan dalam *flowchart* yang merinci dengan jelas bagaimana data diklasifikasikan ke dalam *cluster* yang sesuai dan bagaimana titik-titik pusat *cluster* diperbaharui untuk mendapatkan hasil *cluster* yang paling optimal.

Langkah-langkah dalam algoritma *K-means* dapat diuraikan sebagai berikut. Pertama, penentuan jumlah *cluster* yang ingin dibentuk menjadi langkah awal. Selanjutnya, langkah kedua melibatkan penetapan posisi awal dari *centroid* secara acak, sesuai dengan jumlah *cluster* yang telah ditentukan sebelumnya. Langkah ketiga melibatkan perhitungan jarak antara setiap data dengan *centroid* menggunakan rumus Euclidean distance sesuai dengan persamaan (2) yang telah ditentukan sebelumnya.

Setelah itu, langkah keempat melibatkan pengelompokan data berdasarkan jarak minimum antara setiap data dengan *centroid*. Data-data akan ditempatkan dalam *cluster* yang memiliki *centroid* terdekat dengan data tersebut. Kemudian, pada langkah kelima, dilakukan perhitungan ulang untuk menentukan *centroid* baru pada setiap *cluster*. Hal ini dilakukan dengan mengambil rata-rata dari anggota *cluster* yang telah terbentuk



Gambar 1 Flowchart *K-means*

Terakhir, langkah keenam melibatkan perulangan proses tersebut hingga tidak ada perubahan lagi pada anggota *cluster* atau *centroid*. Proses ini memastikan bahwa *cluster* dan *centroid* yang terbentuk telah mencapai titik konvergensi dan tidak mengalami perubahan yang signifikan. Melalui langkah-langkah yang terperinci ini, algoritma *K-means* mampu mengelompokkan data menjadi *cluster* yang sesuai dengan kesamaan atribut dan mencapai distribusi yang optimal.

Agglomerative Hierarchical Clustering (AHC)

Agglomerative Hierarchical Clustering (AHC) adalah metode *clustering* yang berfokus pada pembentukan hierarki kelompok dengan menggabungkan objek-objek data ke dalam *cluster* yang semakin besar. AHC dimulai dengan setiap objek dianggap sebagai *cluster* tunggal, lalu secara iteratif menggabungkan *cluster-cluster* yang paling mirip berdasarkan jarak antara objek-objeknya. Langkah-langkah AHC melibatkan inisialisasi setiap objek sebagai *cluster* individu, kemudian menghitung jarak antara *cluster-cluster* yang berdekatan dan menggabungkan *cluster* yang memiliki jarak terdekat. Proses ini diulang hingga seluruh objek berada dalam satu *cluster* atau hingga jumlah *cluster* yang diinginkan tercapai [13].

Dalam AHC, penghitungan jarak antara *cluster-cluster* berbasis *single linkage* (jarak terdekat). Dalam *single linkage*, jarak antara dua *cluster* diukur berdasarkan jarak terdekat antara dua objek yang berbeda *cluster* tersebut. Rumus untuk menghitung jarak antara dua *cluster* C1 dan C2 dengan *single linkage* adalah [13].

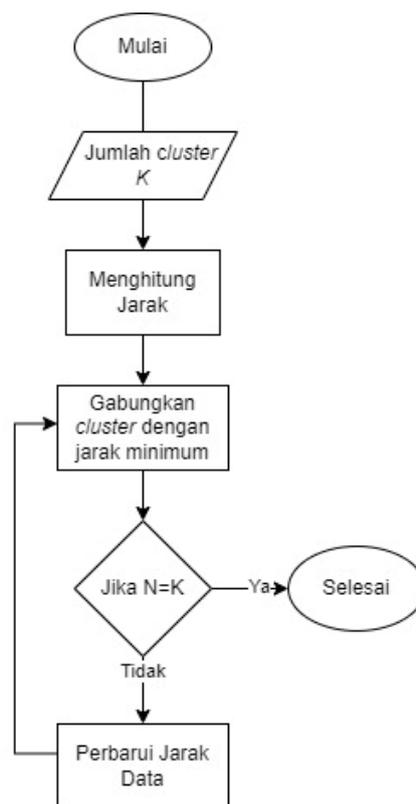
$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \tag{2}$$

Keterangan:

$d_{(UV)W}$: Jarak *linkage* antara *cluster* U dan V saat dikombinasikan dengan *cluster* W.

$\min(d_{UW}, d_{VW})$: nilai minimum dari jarak antara *cluster* U dan W serta jarak antara *cluster* V dan W.

AHC memulai prosesnya dengan menganggap setiap objek sebagai *cluster* individu. Kemudian, *cluster* yang paling mirip ditempatkan bersama dalam tingkat yang lebih tinggi dalam pohon hirarkis, dan proses ini berlanjut hingga semua objek dikelompokkan dalam satu *cluster* besar. Metode AHC menggunakan metrik jarak atau kesamaan antara objek-objek sebagai dasar untuk menentukan kedekatan di antara *cluster* dan objek-objeknya.



Gambar 2 Flowchart AHC

Metode Agglomerative Hierarchical Clustering (AHC) melibatkan serangkaian langkah yang bertujuan untuk mengelompokkan data menjadi *cluster* yang saling berhubungan. Langkah pertama adalah menentukan jumlah *cluster* yang diinginkan, K. Setelah itu, langkah kedua melibatkan perhitungan matriks jarak antara semua pasangan data menggunakan metrik jarak seperti Euclidean distance.

Setelah matriks jarak terbentuk, langkah ketiga menggabungkan dua *cluster* terdekat menjadi satu kelompok data menggunakan metode single linkage. Langkah ini melibatkan pemilihan jarak

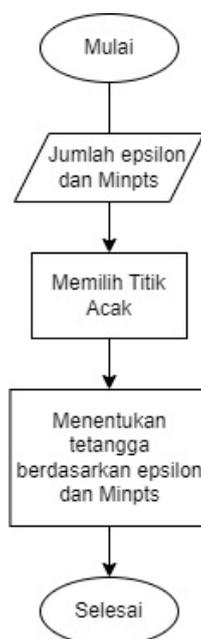
terpendek antara dua *cluster*. Selanjutnya, langkah keempat mengimplikasikan pembaruan matriks jarak untuk mencerminkan jarak antara kelompok baru yang terbentuk dengan kelompok yang masih tersisa.

Proses langkah ketiga dan keempat diulangi hingga terbentuk K *cluster* yang diinginkan. Dengan demikian, metode AHC menghasilkan dendrogram yang menggambarkan secara hierarkis bagaimana data dielompokkan menjadi *cluster*. Proses ini memberikan pemahaman yang lebih mendalam tentang hubungan antara kelompok data dan memungkinkan pemilihan *cluster* yang sesuai dengan interpretasi hasil analisis yang diinginkan.

DBSCAN

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) adalah metode *clustering* yang berfokus pada mengidentifikasi *cluster* berdasarkan kepadatan data di sekitar titik-titik dalam ruang [14]. DBSCAN mengelompokkan titik-titik data yang dekat satu sama lain dalam satu *cluster*, sementara titik-titik yang lebih jauh dari *cluster* dianggap sebagai *noise*. Metode ini memiliki dua parameter utama: ϵ (*epsilon*), yang menunjukkan jarak maksimal antara dua titik agar dianggap tetangga, dan *minPts*, yang mengindikasikan jumlah minimum tetangga dalam radius ϵ agar suatu titik dianggap sebagai titik inti. Proses DBSCAN dimulai dengan memilih titik acak yang belum diunjungi, lalu mengidentifikasi tetangganya dalam radius ϵ . Jika tetangga-tetangga tersebut memenuhi kriteria *minPts*, maka *cluster* baru dibentuk dengan menghubungkan semua titik-titik tetangga. *Cluster* ini kemudian diperluas dengan menambahkan titik-titik tetangga yang memenuhi kriteria, dan proses ini diulang untuk setiap titik yang belum dikunjungi atau belum masuk dalam *cluster* [14].

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) adalah pendekatan *clustering* yang berfokus pada identifikasi *cluster-cluster* dalam data berdasarkan kepadatan objek-objek di sekitar mereka.



Gambar 3 Flowchart DBSCAN

Langkah awal adalah menentukan nilai *epsilon* (radius dari suatu lingkaran) dan juga nilai *Minpts* (jumlah minimum titik yang ada dalam *epsilon*) sebagai parameternya. Setelah nilai-nilai ini ditetapkan, langkah kedua melibatkan pemilihan sebuah titik tengah secara acak dari data yang akan dianalisis.

Kemudian, pada langkah ketiga, algoritma akan menentukan tetangga atau anggota *cluster* berdasarkan jarak dari nilai *epsilon* yang telah ditentukan, serta memeriksa apakah jumlah anggota *cluster* mencapai nilai *Minpts*. Titik-titik yang terletak dalam radius *epsilon* dan memenuhi syarat jumlah anggota lebih dari *Minpts* akan dianggap sebagai bagian dari *cluster* yang sama. Proses ini memungkinkan DBSCAN untuk mendeteksi *cluster* dengan kepadatan yang lebih tinggi, sementara

titik-titik yang kurang padat atau berdiri sendiri akan dianggap sebagai noise atau outlier. Melalui langkah-langkah ini, DBSCAN mampu mengclusterkan data dengan mengakomodasi variasi kepadatan dalam distribusi spasialnya, yang membantu dalam mengidentifikasi cluster yang lebih bermakna dari data yang dianalisis.

Silhouette Coefficient

Silhouette Coefficient adalah metode evaluasi kualitas cluster yang mengukur seberapa baik objek-objek dalam suatu cluster dikelompokkan secara internal, dibandingkan dengan seberapa baik objek-objek tersebut terpisah dari cluster lain. Tujuan dari *Silhouette Coefficient* adalah untuk mengukur sejauh mana objek-objek dalam suatu cluster serupa satu sama lain dan seberapa jauh mereka berbeda dari cluster lainnya [15]. Metode ini memberikan skor antara -1 hingga 1 untuk setiap objek, dimana skor positif menunjukkan bahwa objek tersebut tergolong baik dalam cluster-nya dan berjauhan dari cluster lain, sementara skor negatif menandakan bahwa objek tersebut seharusnya berada dalam cluster lain. Persamaan yang digunakan [13]:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3}$$

Keterangan:

$S(i)$: Nilai *Silhouette Coefficient* untuk objek ke-i dalam cluster.

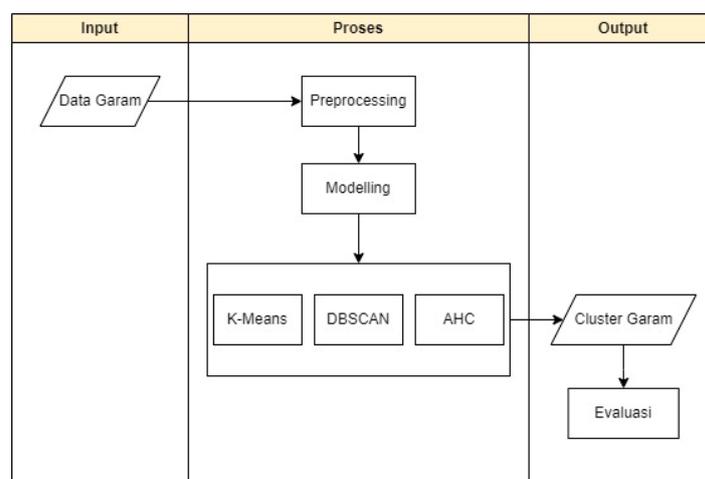
$a(i)$: Rata-rata jarak antara objek ke-i dengan semua objek dalam cluster tempat objek tersebut berada.

$b(i)$: rata-rata jarak antara objek ke-i dan semua objek dalam cluster yang berbeda dengan cluster tempat objek tersebut berada.

HASIL DAN PEMBAHASAN

Arsitektur Sistem

Arsitektur sistem adalah deskripsi urutan proses yang ada dalam sistem untuk memudahkan dalam memahami konsep algoritma. Diagram alur ini menjelaskan algoritma sistem utama mengenai proses clustering pada data garam. Gambaran umum diagram dapat dilihat pada Gambar.1



Gambar 4 Arsitektur Sistem

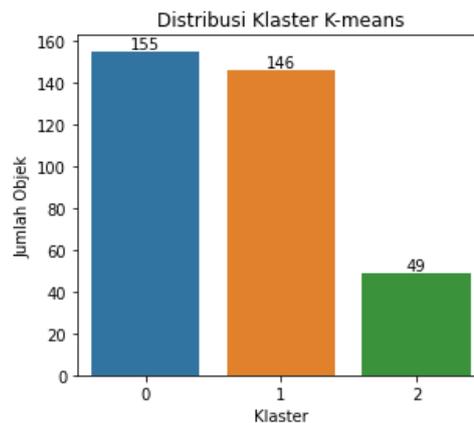
Sistem yang dibangun memiliki 3 tahapan proses, yaitu Input, Proses dan Output.

1. Input data garam yang akan digunakan pada proses penelitian,

2. Tahap *preprocessing*, yaitu seleksi fitur dan normalisasi data dengan *minmax normalization*.
3. Data yang sudah dilakukan *preprocessing* akan dilakukan *modelling* menggunakan algoritma *K-means*, DBSCAN, dan AHC.
4. Setelah *cluster* terbentuk, akan dihitung evaluasinya menggunakan *Silhouette Coefficient*.

K-Means

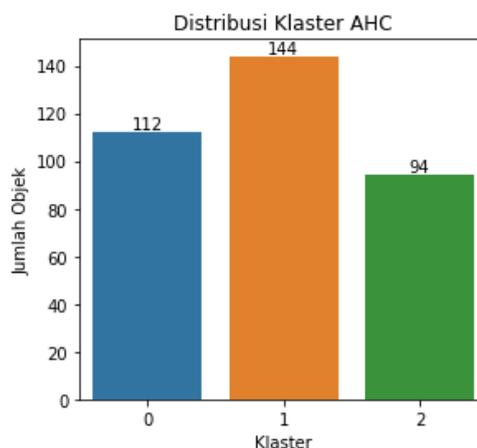
Pada percobaan ini, ditentukan jumlah *cluster* sebesar 3. Sehingga dapat mengidentifikasi kelompok-kelompok berbeda dalam suatu data. Anggota tiap *cluster* hasil dari algoritma *K-means* masing-masing adalah, *cluster* 1 sebanyak 155, *cluster* 2 sebanyak 146 dan *cluster* 3 sebanyak 49 obyek. Pengelompokkan ini memungkinkan untuk mengetahui suatu karakteristik tiap kelompoknya. Grafik jumlah tiap *cluster* nya dapat dilihat pada Gambar



Gambar 5 Distribusi *Cluster K-means*

AHC (Agglomerative Hierarchical Clustering)

Pada penelitian ini, AHC menghasilkan tiga *cluster* dengan jumlah objek yang berbeda. *Cluster* pertama memiliki 112 objek, *cluster* kedua memiliki 144 objek, dan *cluster* ketiga memiliki 94 objek. Interpretasi dari hasil ini dapat memberikan wawasan tentang bagaimana data terorganisir dan bagaimana objek-objek tersebut berkelompok berdasarkan kesamaan atribut.

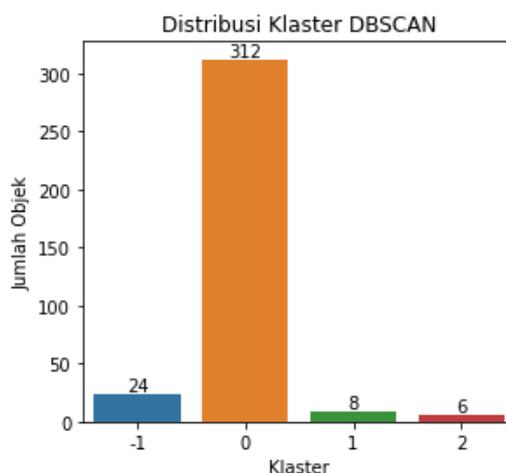


Gambar 6 Distribusi *Cluster AHC*

DBSCAN

Dalam skenario ini, hasil DBSCAN menghasilkan empat *cluster* dengan ukuran masing-masing 312, 8, dan 6 objek. DBSCAN bekerja dengan cara menghubungkan objek-objek yang memiliki kepadatan yang cukup dekat dan meng-kelompokkannya ke dalam satu *cluster*. Selain itu, objek-objek yang memiliki kepadatan rendah atau terisolasi dianggap sebagai *noise* dan tidak termasuk dalam *cluster*.

Cluster dengan label-1 mengindikasikan bahwa objek tersebut tidak dapat distribusikan ke *cluster* manapun karena tidak memenuhi kriteria kepadatan atau tidak memiliki tetangga yang cukup dalam jarak ϵ . Ini memungkinkan DBSCAN untuk mengidentifikasi dan mengabaikan objek-objek yang mungkin tidak bermakna dalam analisis *cluster*, serta mengisolasi mereka dari *cluster* yang signifikan dalam data.



Gambar.7 Distribusi Cluster DBSCAN

Berdasarkan pengujian dari tiga metode *cluster* K-means, AHC, dan DBSCAN menggunakan pengukuran *silhouette coefficient* dapat dilihat pada tabel 1.

Tabel 1. Perbandingan *silhouette coefficient*

No	Metode	<i>silhouette coefficient</i>
1.	K-means	0.345
2.	AHC	0.32
3.	DBSCAN	0.5

Nilai *cluster* yang tertinggi diperoleh metode DBSCAN sebesar 0.5 sehingga struktur *silhouette coefficient* lebih kuat dibanding dengan K-means dan AHC dengan nilai berturut-turut 0.345 dan 0.32. Hal ini dikarenakan jika nilai *silhouette coefficient* mendekati 1 maka dapat dikatakan struktur clusternya lebih kuat.

KESIMPULAN

Kesimpulan yang didapatkan dari beberapa uji coba yang dilakukan dapat disimpulkan bahwa algoritma *clustering K-means* dan AHC harus menentukan nilai K untuk penentuan *cluster*. Sehingga perlu diperhatikan untuk pemilihan nilai K. kemudian penentuan *epsilon* dan Minpts sangat berpengaruh pada Algoritma DBSCAN. *Epsilon* dan Minpts yang terlalu besar mengakibatkan semua data dianggap tetangga oleh algoritma. Dari jumlah data sebanyak 350 dengan 7 fitur yang berasal dari PT. Garam Persero yang di kelompokkan menggunakan tiga metode dilakukan pengujian kualitas *clustering* menggunakan *silhouette coefficient* yang menghasilkan nilai 0.345 untuk metode *K-means*,

0.32 untuk metode AHC dan 0.5 untuk metode DBSCAN. Saran pengembangan dari penelitian ini supaya mendapatkan hasil cluster yang maksimal adalah dengan melakukan proses penggalian variasi variabel data yang berbeda.

UCAPAN TERIMA KASIH

Terima kasih peneliti ucapkan kepada LPPM Universitas Trunojoyo Madura yang telah menyediakan dana, sarana dan prasarana yang dapat menunjang penelitian ini berlangsung. Utamanya kepada Fakultas teknik, Jurusan teknik Informatika dimana peneliti bernaung. Laboratorium Multimedia dan jaringan dimana penelitian ini dilaksanakan.

DAFTAR RUJUKAN

- [1] Y. Kustiyahningsih, E. Rahmanita, and N. Kholifah, "Salt Farmer measurement performance system facing Covid-19 pandemic used interval type-2 FAHP Method," in *Journal of Physics: Conference Series*, 2022, vol. 2193, no. 1, p. 12012, doi: 10.1088/1742-6596/2193/1/012012.
- [2] Y. Kustiyahningsih, E. Rahmanita, E. M. S. Rochman, A. Amalina, and M. H. A. Sobri, "Decision Support System of Salt Points Grouping Using K-Means Method," in *Journal of Physics: Conference Series*, 2022, vol. 2406, no. 1, p. 12022, doi: 10.1088/1742-6596/2406/1/012022.
- [3] M. Kurdi, "Beyond Salt Industries and Environment in Sumenep: Effective Partnership for People Welfare," in *IOP Conference Series: Earth and Environmental Science*, 2020, vol. 469, no. 1, p. 12085, doi: 10.1088/1755-1315/469/1/012085.
- [4] G. Liu, F. Ji, W. Sun, and L. Sun, "Optimization design of short-circuit test platform for the distribution network of integrated power system based on improved K-means clustering," *Energy Reports*, vol. 9, pp. 716–726, 2023, doi: 10.1016/j.egyr.2023.04.319.
- [5] D. Das, P. Kayal, and M. Maiti, "A K-means clustering model for analyzing the Bitcoin extreme value returns," *Decis. Anal. J.*, vol. 6, p. 100152, 2023, doi: 10.1016/j.dajour.2022.100152.
- [6] W. Kwedlo and M. Łubowicz, "Accelerated K-means algorithms for low-dimensional data on parallel shared-memory systems," *IEEE Access*, vol. 9, pp. 74286–74301, 2021, doi: 10.1109/ACCESS.2021.3080821.
- [7] E. M. S. Rochman and H. Suprajitno, "Comparison of clustering in tuberculosis using fuzzy c-means and k-means methods," *Commun. Math. Biol. Neurosci.*, vol. 2022, p. Article-ID, 2022, doi: 10.28919/cmbn/7335.
- [8] R. M. F. Lubis, J.-P. Huang, P.-C. Wang, N. Damanik, A. C. Sitepu, and C. D. Simanullang, "K-Means and AHC Methods for Classifying Crime Victims by Indonesian Provinces: A Comparative Analysis," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 295–307, 2023, doi: 10.47065/bits.v5i1.3630.
- [9] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijis.v4i1.73.
- [10] A. Pandey and A. Jain, "Comparative analysis of KNN algorithm using various normalization techniques," *Int. J. Comput. Netw. Inf. Secur.*, vol. 10, no. 11, p. 36, 2017, doi:

10.5815/ijcnis.2017.11.04.

- [11] M. Faisal and E. M. Zamzami, “Comparative analysis of inter-centroid K-Means performance using euclidean distance, canberra distance and manhattan distance,” in *Journal of Physics: Conference Series*, 2020, vol. 1566, no. 1, p. 12112, doi: 10.1088/1742-6596/1566/1/012112.
- [12] A. R. Lubis and M. Lubis, “Optimization of distance formula in K-Nearest Neighbor method,” *Bull. Electr. Eng. Informatics*, vol. 9, no. 1, pp. 326–338, 2020, doi: 10.11591/eei.v9i1.1464.
- [13] E. K. A. Mala, S. Rochman, I. K. A. O. Suzanti, R. Jannah, B. K. Khotimah, and A. Rachmad, “A combination of algorithm agglomerative hierarchical cluster (AHC) and K-means for clustering tourism in Madura-Indonesia,” *J. Math. Comput. Sci.*, pp. 1–19, 2022, doi: 10.28919/jmcs/7086.
- [14] S. Wibisono, M. T. Anwar, A. Supriyanto, and I. H. A. Amin, “Multivariate weather anomaly detection using DBSCAN clustering algorithm,” *J. Phys. Conf. Ser.*, vol. 1869, no. 1, 2021, doi: 10.1088/1742-6596/1869/1/012077.
- [15] X. Wang and Y. Xu, “An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 569, no. 5, 2019, doi: 10.1088/1757-899X/569/5/052024.