

**LAPORAN TUGAS BESAR DATA MINING**  
**“SEGMENTASI PELANGGAN PRODUK SARI ROTI DENGAN DBSCAN**  
***CLUSTERING ANALYSIS*”**



**Dosen Pengampu:**

1. Aina Hubby Azira, M.Eng.
2. Dwi Welly Sukma Nirad, M.T.

**Disusun Oleh Kelompok 1:**

Muhammad Nouval Habibie	2211521020
Nabil Rizki Navisa	2211522018
Ilham Nofaldi	2211522028
Mustafa Fathur Rahman	2211522036
Syauqi Nabiih Marwa	2211523012

**DEPARTEMEN SISTEM INFORMASI**  
**FAKULTAS TEKNOLOGI INFORMASI**  
**UNIVERSITAS ANDALAS**

**2024**

## KATA PENGANTAR

Segala puji bagi Allah SWT yang telah memberikan kami kemudahan sehingga kami dapat menyelesaikan laporan ini tepat waktu. Tanpa pertolonganNya tentunya kami tidak akan sanggup untuk menyelesaikan laporan ini dengan baik. Shalawat serta salam semoga terlimpah curahkan kepada baginda tercinta kita yaitu Nabi Muhammad SAW yang kita nanti-nantikan syafa'atnya di akhirat nanti.

Penulis mengucapkan syukur kepada Allah SWT atas limpahan nikmat sehat-Nya, baik itu berupa sehat fisik maupun akal pikiran, sehinggann penulis mampu untuk menyelesaikan pembuatan laporan tugas besar dari mata kuliah Data Mining dengan judul “Segmentasi Pelanggan Produk Sari Roti dengan Algoritma DBSCAN *Clustering Analysis*”.

Dalam penyusunan laporan tugas besar ini, penulis banyak memperoleh bantuan dan masukan dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Tuhan yang Maha Esa
2. Ibu Aina Hubby Azira, M.Eng. dan Dwi Welly Sukma Nirad, M.T. selaku dosen pengampu
3. Uda Arif, rekan-rekan mahasiswa, dan sumber lainnya yang menjadi referensi penulis dalam menyusun laporan tugas besar ini.

Penulis tentu menyadari bahwa laporan ini masih jauh dari kata sempurna dan masih banyak terdapat kesalahan serta kekurangan di dalamnya. Untuk itu, penulis mengharapkan kritik serta saran dari pembaca laporan ini, supaya laporan ini nantinya dapat menjadi laporan yang lebih baik lagi. Demikian, dan apabila terdapat banyak kesalahan pada makalah ini penulis mohon maaf yang sebesar-besarnya. Semoga laporan ini dapat berguna baik untuk pembaca maupun penulis sendiri.

Padang, 22 Mei 2024

## DAFTAR ISI

KATA PENGANTAR.....	2
DAFTAR ISI.....	3
DAFTAR GAMBAR.....	5
DAFTAR TABEL.....	6
DAFTAR CODE.....	7
BAB I.....	8
PENDAHULUAN .....	8
1.1    Latar Belakang.....	8
1.2    Rumusan Masalah .....	9
1.3    Batasan Masalah .....	10
1.4    Tujuan Penelitian .....	10
1.5    Manfaat Penelitian.....	11
BAB II.....	12
METODOLOGI.....	12
2.1    Distributor .....	12
2.2    Segmentasi Pelanggan.....	13
2.3    Data Mining.....	14
2.4    Clustering .....	15
2.5    DBSCAN ( <i>Density-Based Spatial Clustering of Applications with Noise</i> ) .....	16
2.6    Tahapan Data Mining.....	17
BAB III.....	20
HASIL DAN PEMBAHASAN .....	20
3.1    Dataset.....	20
3.2    Implementasi .....	23
3.2.1    Preprocessing Data .....	23
3.3    Penerapan Algoritma DBSCAN dengan perhitungan manual. ....	26
3.4    Proses Clustering dengan Algoritma DBSCAN.....	31
BAB IV.....	38
PENUTUP.....	38
4.1    Kesimpulan.....	38
4.2    Saran .....	38

<b>DAFTAR PUSTAKA .....</b>	<b>40</b>
-----------------------------	-----------

## **DAFTAR GAMBAR**

Gambar 1 Tahapan Data Mining .....	18
Gambar 2 visualisasi cluster menggunakan 3d plot.....	34
Gambar 3 jumlah anggota tiap cluster .....	36

## DAFTAR TABEL

Tabel 1 Dataset.....	20
Tabel 2 Dataset.....	20
Tabel 3 Dataset.....	20
Tabel 4 Dataset.....	24
Tabel 5 transformasi dan normalisasi data .....	26
Tabel 6 hasil perhitungan iterasi 1 .....	28
Tabel 7 titik terseleksi dengan $\text{eps} = 5$ .....	28
Tabel 8 iterasi 2 .....	29
Tabel 9 jarak titik $g$ .....	29
Tabel 10 iterasi 3 .....	30
Tabel 11 jarak ke titik $i$ .....	30
Tabel 12 cluster sementara dari setiap iterasi.....	30

## DAFTAR CODE

Code 1 import library dan membaca dataset.....	31
Code 2 info dataset.....	32
Code 3 deklarasi dbscan.....	32
Code 4 visualisasi cluster.....	33
Code 5 melihat list cluster.....	35
Code 6 jumlah anggota tiap cluster.....	36

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Data mining saat ini menjadi salah satu teknik yang sering digunakan untuk mengolah data menjadi informasi bermanfaat. Dalam era digital, data menjadi aset yang sangat berharga bagi perusahaan bisnis di berbagai sektor. Teknik data mining memungkinkan perusahaan untuk menemukan pola tersembunyi dan mendapatkan wawasan yang berguna dari sebuah data yang besar dan kompleks, salah satu sektor yang dapat menerapkan teknik data mining adalah perusahaan distribusi dengan produk berupa makanan Sari Roti.

Distribusi makanan adalah proses penting yang melibatkan pengiriman produk dari produsen ke berbagai titik penjualan seperti kedai dan toko. Perusahaan distribusi berfokus pada memastikan bahwa produk, dalam hal ini Sari Roti, didistribusikan dengan efisien dan tepat waktu ke berbagai kedai. Proses bisnis ini mencakup pengelolaan logistik, pengiriman produk, dan pencatatan data penjualan di masing-masing kedai. Keberhasilan dalam distribusi sangat tergantung pada kemampuan perusahaan dalam mengelola rantai pasokan dan menjaga ketersediaan produk di seluruh jaringan penjualannya.

Perusahaan distribusi Sari Roti menghadapi tantangan dalam mengelompokkan pelanggan berdasarkan pola pembelian mereka, termasuk waktu dan jenis produk yang dibeli. Misalnya, pada gerai tertentu, jenis produk Sari Roti A sangat laku di hari Senin, sementara itu, produk Sari Roti B tidak terjual sama sekali di hari yang sama. Meskipun data penjualan di setiap kedai dicatat, tanpa segmentasi yang tepat, perusahaan kesulitan mengidentifikasi tren dan membuat prediksi permintaan yang akurat. Hal ini dapat mengakibatkan masalah seperti kelebihan stok atau kekurangan stok, yang berdampak negatif pada efisiensi operasional dan kepuasan pelanggan.

Untuk mengatasi tantangan tersebut, penelitian ini mengusulkan penggunaan salah satu metode pada data mining, yaitu *clustering*. *Clustering* adalah teknik dalam data mining yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki kesamaan tertentu. Dalam konteks distribusi Sari Roti, ini berarti mengelompokkan toko-toko berdasarkan pola pembelian pelanggan mereka. Misalnya, beberapa toko mungkin lebih



cenderung menjual Sari Roti A pada hari tertentu, sementara toko lainnya mungkin lebih aktif menjual Sari Roti B di hari yang sama. Dalam distribusi makanan, memahami pola pembelian pelanggan adalah kunci untuk mengoptimalkan distribusi dan manajemen stok. *Clustering* merupakan solusi yang tepat karena tidak memerlukan label data sebelumnya dan mampu mengungkap pola pembelian yang mungkin tidak terlihat secara langsung.

Metode *clustering* sendiri memiliki beberapa jenis. Untuk penelitian ini, penulis memilih menggunakan algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). DBSCAN adalah algoritma yang mengelompokkan data berdasarkan kepadatan titik-titik data dalam ruang. Dalam konteks ini, algoritma DBSCAN akan digunakan untuk mengelompokkan kedai berdasarkan waktu pembelian dan jenis produk Sari Roti yang dibeli. DBSCAN dipilih sebagai solusi konkret karena memiliki keunggulan dalam mengidentifikasi *cluster* dengan bentuk arbitrer dan mampu menangani data dengan *noise* (data yang tidak termasuk dalam *cluster* manapun).

Penelitian ini dilakukan untuk mengatasi tantangan distribusi dengan menggunakan metode DBSCAN untuk segmentasi pelanggan. Hal ini akan membantu perusahaan untuk mengetahui bagaimana segmentasi pelanggan berdasarkan waktu dan jenis produk yang dibeli pada setiap toko atau gerai. Dengan hasil yang didapatkan dari perhitungan tersebut, perusahaan distribusi Sari Roti akan mendapatkan sebuah pengetahuan berupa informasi yang akan sangat berguna jika di kemudian hari pihak perusahaan distribusi ingin membuat prediksi permintaan di setiap toko atau gerai sehingga dapat meningkatkan efektivitas dalam pengelolaan stok.

## 1.2 Rumusan Masalah

Dari latar belakang yang dijelaskan, beberapa rumusan masalah yang dapat diidentifikasi adalah:

1. Mengapa algoritma DBSCAN *clustering* dipilih sebagai solusi konkret untuk segmentasi pelanggan Sari Roti?

2. Bagaimana cara mengelompokkan pelanggan Sari Roti berdasarkan pola pembelian mereka, terutama dalam hal waktu dan jenis produk yang dibeli dengan algoritma DBSCAN *clustering*
3. Bagaimana hasil *cluster* segmentasi pelanggan dapat membantu perusahaan distribusi Sari Roti dalam melihat tren atau pola penjualan produk di setiap outlet?

### 1.3 Batasan Masalah

Dengan membatasi penelitian pada "Segmentasi Pelanggan Produk Sari Roti Dengan Algoritma DBSCAN Clustering Analysis," fokusnya adalah:

1. Segmentasi Pelanggan Sari Roti: Penelitian akan memusatkan perhatian pada pengelompokkan pelanggan Sari Roti berdasarkan pola pembelian mereka.
2. Penggunaan Algoritma DBSCAN: Metode utama yang digunakan adalah algoritma DBSCAN karena kemampuannya dalam mengidentifikasi kelompok data berkerumun dan menangani data dengan keberagaman kepadatan.
3. Analisis Pola Pembelian: Fokus analisis adalah pola pembelian pelanggan, terutama waktu pembelian dan jenis produk yang dibeli.
4. Data yang Digunakan: Data berasal dari catatan penjualan Sari Roti di berbagai outlet selama periode tertentu.

Dengan pembatasan ini, diharapkan penelitian akan lebih efektif dan efisien dalam mencapai tujuan segmentasi pelanggan Sari Roti.

### 1.4 Tujuan Penelitian

Dari rumusan masalah yang telah diuraikan sebelumnya, maka tujuan penelitian ini adalah sebagai berikut:

1. Untuk mengetahui mengapa algoritma DBSCAN *clustering* dipilih sebagai solusi konkret untuk segmentasi pelanggan Sari Roti.

2. Untuk mengimplementasikan algoritma DBSCAN *clustering* dalam mengelompokkan pelanggan Sari Roti berdasarkan pola pembelian mereka, terutama dalam hal waktu dan jenis produk yang dibeli,
3. Untuk mengevaluasi bagaimana hasil dari segmentasi pelanggan dengan menggunakan algoritma k-means *clustering* dapat membantu perusahaan distribusi Sari Roti dalam melihat tren atau pola penjualan produk di setiap outlet,

### **1.5 Manfaat Penelitian**

Apabila tujuan dari penelitian berhasil tercapai maka akan didapatkan beberapa manfaat dari penelitian ini, yaitu:

1. Memberikan wawasan yang lebih dalam tentang pola pembelian pelanggan Sari Roti, sehingga perusahaan merencanakan pemasaran yang lebih efektif.
2. Membantu perusahaan dalam mengoptimalkan manajemen stok dan distribusi produk dengan memprediksi permintaan pelanggan di setiap gerai atau toko.
3. Menjadi landasan untuk penelitian lanjutan dalam bidang segmentasi pelanggan dan manajemen rantai pasokan, serta dapat diterapkan pada industri distribusi makanan dan minuman lainnya.

## **BAB II**

### **METODOLOGI**

#### **2.1 Distributor**

Kata "distributor" berasal dari kata "distribusi." Distribusi adalah proses penyebaran atau pembagian barang atau jasa kepada beberapa orang atau ke beberapa tempat. Menurut Kamus Besar Bahasa Indonesia (KBBI), distribusi adalah "penyebaran atau pembagian kepada beberapa orang atau ke beberapa tempat" (KBBI) [6]. Dalam konteks bisnis, distribusi mencakup seluruh aktivitas yang terlibat dalam memindahkan produk dari produsen ke konsumen akhir. Proses distribusi melibatkan berbagai elemen seperti pergudangan, pengangkutan, penanganan produk, dan pengelolaan saluran distribusi yang tepat untuk memastikan barang sampai kepada konsumen dengan efisien dan efektif.

Distributor adalah salah satu komponen utama dalam proses distribusi. Menurut KBBI, distributor adalah "orang atau perusahaan yang bertindak sebagai perantara untuk menyalurkan barang dagangan dari produsen ke pengecer atau pemakai" (KBBI) [6]. Distributor memainkan peran penting dalam rantai pasokan dengan membantu produk mencapai konsumen akhir secara efisien.

Dalam konteks distribusi produk Sari Roti, distributor Sari Roti bertugas memastikan produk roti yang diproduksi oleh PT Nippon Indosari Corpindo Tbk. dapat sampai ke tangan konsumen di berbagai daerah di Indonesia. Distributor Sari Roti biasanya membeli produk dalam jumlah besar dari produsen dan kemudian menjualnya kembali dalam jumlah yang lebih kecil kepada pengecer atau langsung kepada konsumen.

Distributor Sari Roti juga bertanggung jawab untuk mengelola inventaris, penyimpanan, dan pengiriman barang. Mereka sering kali memiliki fasilitas penyimpanan yang besar dan sistem logistik yang efisien untuk memastikan produk dapat dikirim dengan cepat dan dalam kondisi baik. Selain itu, distributor Sari Roti juga dapat memberikan dukungan pemasaran dan penjualan kepada pengecer untuk membantu mempromosikan produk kepada konsumen.

Dengan peran ini, distributor membantu memperluas jangkauan pasar bagi Sari Roti, meningkatkan ketersediaan produk, dan memastikan bahwa produk dapat diakses oleh

konsumen di berbagai lokasi. Efektivitas distributor dalam menjalankan tugasnya sangat mempengaruhi keberhasilan strategi distribusi perusahaan secara keseluruhan.

## **2.2 Segmentasi Pelanggan**

Segmentasi pelanggan adalah proses membagi pasar menjadi kelompok-kelompok yang berbeda berdasarkan kebutuhan atau karakteristik serupa. Menurut Kotler dan Keller (2016), segmentasi pasar adalah tindakan membagi pasar menjadi beberapa kelompok pembeli yang berbeda berdasarkan kebutuhan, karakteristik, atau perilaku yang berbeda yang mungkin membutuhkan produk atau bauran pemasaran yang berbeda [7]. Dalam konteks ini, pelanggan yang dimaksud adalah gerai-gerai yang menjadi target distribusi Sari Roti.

Dalam penelitian ini, segmentasi pelanggan dilakukan berdasarkan volume dropping produk ke toko-toko. Volume dropping mengacu pada jumlah produk yang dikirim ke setiap toko selama periode waktu tertentu. Dengan mengelompokkan toko-toko berdasarkan volume dropping, Sari Roti dapat mengidentifikasi pola distribusi dan menyesuaikan strategi distribusinya untuk memenuhi kebutuhan masing-masing segmen toko.

Segmentasi berdasarkan volume dropping produk membantu Sari Roti untuk memahami dan mengelola variasi permintaan di berbagai toko. Beberapa toko mungkin menerima volume produk yang lebih besar karena permintaan yang tinggi, sementara toko lain mungkin menerima volume yang lebih kecil. Dengan melakukan segmentasi ini, perusahaan dapat mengoptimalkan rantai pasokan, mengurangi biaya distribusi, dan meningkatkan kepuasan pelanggan dengan memastikan bahwa setiap toko memiliki stok yang sesuai dengan permintaan mereka.

Dalam konteks ini, metode clustering seperti algoritma DBSCAN dapat digunakan untuk mengelompokkan toko-toko berdasarkan volume dropping mereka. DBSCAN adalah algoritma clustering yang mengelompokkan data berdasarkan kepadatan titik-titik data dalam ruang. Ini memungkinkan identifikasi cluster dengan bentuk yang arbitrer dan mampu menangani data dengan noise (data yang tidak termasuk dalam cluster manapun). Dengan menggunakan DBSCAN, Sari Roti dapat mengelompokkan toko-toko yang memiliki pola

volume dropping serupa, sehingga memudahkan perusahaan untuk mengelola dan merencanakan distribusi secara lebih efektif.

Penelitian ini dilakukan untuk mengatasi tantangan distribusi dengan menggunakan metode DBSCAN untuk segmentasi pelanggan. Hal ini akan membantu perusahaan untuk mengetahui bagaimana segmentasi pelanggan berdasarkan volume dropping produk yang dibeli pada setiap gerai. Dengan hasil yang didapatkan dari perhitungan tersebut, perusahaan distribusi Sari Roti akan mendapatkan sebuah pengetahuan berupa informasi yang akan sangat berguna jika di kemudian hari pihak perusahaan distribusi ingin membuat prediksi permintaan di setiap gerai sehingga dapat meningkatkan efektivitas dalam pengelolaan stok.

## **2.3 Data Mining**

Data mining adalah proses ekstraksi pola atau informasi bermanfaat dari sekumpulan data yang besar melalui berbagai teknik dan algoritma tertentu [1]. Data mining merupakan bagian dari Knowledge Discovery in Database (KDD), yang merupakan proses menggali informasi tersembunyi dalam suatu basis data yang besar [2].

Konsep dasar data mining berhubungan dengan bidang ilmu seperti statistika, machine learning, manajemen basis data, dan information retrieval. Tujuan utama data mining adalah menemukan pola atau model dari sekumpulan data yang besar dengan menggunakan teknik seperti klasifikasi, clustering, asosiasi, dan lain-lain [3].

Secara umum, tujuan dari data mining dapat dirangkum sebagai berikut [4]:

- 1) Deskripsi: Memberikan gambaran umum atau mendeskripsikan pola dan kecenderungan yang terdapat dalam data.
- 2) Estimasi: Memperkirakan atau memprediksi nilai numerik dari suatu objek data berdasarkan pola dalam data.
- 3) Prediksi: Memprediksi nilai masa depan atau keanggotaan dalam suatu kelas atau kelompok berdasarkan pola dalam data.
- 4) Klasifikasi: Mengklasifikasikan objek data ke dalam kategori atau kelas tertentu berdasarkan pola dalam data.

- 5) Clustering: Mengelompokkan objek data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan karakteristiknya.
- 6) Asosiasi: Menemukan pola atau aturan asosiasi antara kombinasi item dalam sekumpulan data.

Data mining memiliki aplikasi di berbagai bidang seperti pemasaran, keuangan, manufaktur, telekomunikasi, kesehatan, dan lain-lain [5].

## ***2.4 Clustering***

Clustering merupakan salah satu teknik utama dalam data mining yang bertujuan untuk mengelompokkan objek data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan karakteristiknya [1]. Clustering termasuk dalam metode pembelajaran tanpa pengawasan (unsupervised learning), di mana tidak ada label atau kelas yang ditentukan sebelumnya untuk objek data. Clustering berupaya menemukan struktur atau pola pengelompokan yang terdapat dalam data secara alami.

Tujuan utama dari clustering adalah untuk mengidentifikasi kelompok-kelompok objek data yang memiliki kemiripan karakteristik di dalam kelompok yang sama (intra-cluster similarity) dan memiliki perbedaan karakteristik dengan objek di kelompok lain (inter-cluster dissimilarity) [2]. Dengan kata lain, clustering mencoba menemukan kelompok-kelompok objek data yang kompak dan terpisah.

Algoritma clustering memiliki beberapa karakteristik utama, yaitu [8]:

- 1) Ukuran Kemiripan: Metode untuk mengukur kemiripan atau ketidakmiripan (jarak) antara dua objek data.
- 2) Kriteria Pengelompokan: Kriteria yang digunakan untuk menentukan pemisahan dan penggabungan cluster.
- 3) Pengelompokan Objek: Proses pengelompokan objek data ke dalam cluster-cluster berdasarkan kriteria tertentu.
- 4) Representasi Cluster: Metode untuk merepresentasikan cluster, seperti menggunakan centroid, distribusi probabilitas, atau objek representatif.

Beberapa algoritma clustering yang populer, antara lain [9]:

- 1) K-Means: Algoritma yang membagi data ke dalam k cluster dengan meminimalkan jumlah kuadrat jarak dari setiap objek data ke centroid terdekat.
- 2) Hierarchical Clustering: Algoritma yang membangun hierarki cluster dengan penggabungan atau pemisahan secara bertingkat.
- 3) DBSCAN: Algoritma yang mengelompokkan objek data berdasarkan kepadatan wilayah dan dapat mendeteksi outlier.
- 4) Mean-Shift: Algoritma yang mencari pusat-pusat cluster secara iteratif berdasarkan pergeseran rata-rata (mean shift).

Clustering memiliki aplikasi di berbagai bidang, seperti segmentasi pelanggan, analisis genom, pengelompokan dokumen, dan deteksi anomali [10].

## **2.5 DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*)**

DBSCAN adalah algoritma clustering berbasis kepadatan yang mampu mendeteksi cluster dengan bentuk sembarang dan mengidentifikasi noise atau outlier dalam data [11]. Algoritma ini membangun cluster berdasarkan kepadatan area lokal dari objek data. Konsep utama dari DBSCAN adalah membentuk cluster dari objek data yang saling berdekatan dan menolak objek yang terletak di area dengan kepadatan rendah (dianggap sebagai noise).

DBSCAN menggunakan dua parameter utama, yaitu:

- 1) Eps ( $\epsilon$ ): Jarak maksimum antara dua objek untuk dianggap sebagai tetangga.
- 2) MinPts: Jumlah minimum objek tetangga dalam jarak Eps yang dibutuhkan untuk membentuk sebuah cluster.

Algoritma DBSCAN bekerja dengan cara mengunjungi setiap objek data dalam dataset dan melakukan pengelompokan berdasarkan kepadatan area lokal. Proses pengelompokan dilakukan sebagai berikut [12]:



- 1) Objek data dikelompokkan ke dalam tiga kategori: core point, border point, dan noise point.
- 2) Core point adalah objek yang memiliki setidaknya MinPts objek tetangga dalam radius Eps.
- 3) Border point adalah objek yang bukan core point, tetapi berdekatan dengan core point.
- 4) Noise point adalah objek yang tidak termasuk core point atau border point.
- 5) Cluster dibentuk dari core point dan border point yang saling terhubung.

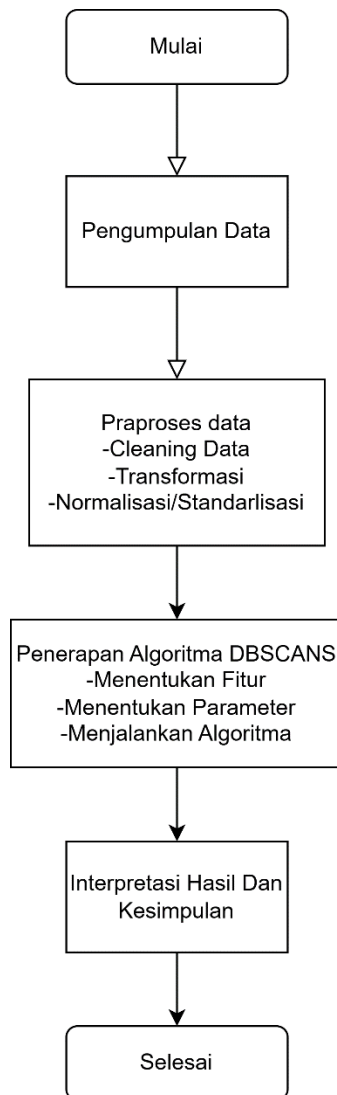
Kelebihan utama DBSCAN antara lain [1]:

- 1) Dapat mendeteksi cluster dengan bentuk sembarang
- 2) Dapat mengidentifikasi noise atau outlier dalam data
- 3) Tidak memerlukan informasi jumlah cluster di awal
- 4) Relatif tidak terpengaruh oleh data yang memiliki kepadatan yang berbeda

DBSCAN cocok digunakan untuk data dengan kepadatan yang bervariasi dan dapat menangani cluster dengan bentuk tidak beraturan atau *arbitrary shape*.

## 2.6 Tahapan Data Mining

Data mining terdiri dari serangkaian tahapan yang sistematis untuk mengekstraksi informasi atau pengetahuan dari data. Tahapan-tahapan ini saling terkait dan membentuk suatu siklus iteratif. Berikut adalah penjelasan terperinci untuk setiap tahapan dalam proses data mining:



Gambar 1 Tahapan Data Mining

#### 1) Pengumpulan Data

Langkah awal yang dilakukan adalah mengumpulkan data penjualan Sari Roti dari berbagai gerai atau toko. Data ini mencakup informasi seperti tanggal pembelian, jenis produk yang dibeli, dan kuantitas pembelian.

#### 2) Praproses Data

Sebelum melakukan *clustering*, praproses data perlu dilakukan untuk memastikan kualitas data yang baik. Tahapan ini meliputi menangani data yang hilang atau tidak lengkap,

menghapus outlier atau data yang menyimpang, serta melakukan normalisasi atau standarisasi data (jika diperlukan).

3) Memilih Fitur yang Relevan

Pilih kolom atau fitur yang dibutuhkan untuk mendapatkan informasi yang diinginkan.

4) Menentukan Parameter DBSCAN

DBSCAN memerlukan dua parameter utama, yaitu *eps* (epsilon) dan *min\_samples*. *Eps* merupakan jarak maksimum antara dua titik data untuk dianggap sebagai tetangga, sedangkan *min\_samples* adalah jumlah minimum titik data dalam suatu tetangga untuk membentuk sebuah *cluster*. Pemilihan nilai parameter ini sangat penting dan dapat mempengaruhi hasil *clustering*.

5) Menjalankan Algoritma DBSCAN

Setelah menentukan parameter, algoritma DBSCAN dijalankan pada data. Algoritma ini akan mengelompokkan data berdasarkan kepadatan (density) dan memberi label *cluster* pada setiap titik data. Titik data yang tidak termasuk dalam *cluster* akan dianggap sebagai noise.

6) Interpretasi Hasil *Clustering*

Hasil *clustering* diinterpretasikan dengan melihat setiap *cluster* yang terbentuk. Setiap *cluster* mewakili segmen pelanggan dengan pola pembelian yang mirip dalam hal waktu dan jenis produk yang dibeli. Visualisasi data, seperti scatter plot atau grafik, dapat digunakan untuk memudahkan interpretasi.

7) Evaluasi dan Penarikan Kesimpulan

Langkah terakhir adalah mengevaluasi hasil *clustering* dan menarik kesimpulan yang relevan dengan tujuan penelitian. Misalnya, mengidentifikasi segmen pelanggan mana yang paling banyak membeli produk tertentu pada hari tertentu, atau menganalisis pola pembelian yang berbeda di antara segmen pelanggan.

Hasil *clustering* DBSCAN ini dapat membantu perusahaan distribusi Sari Roti dalam memahami pola pembelian pelanggan, sehingga mereka dapat mengoptimalkan manajemen stok, distribusi produk, dan strategi pemasaran yang sesuai dengan setiap segmen pelanggan.

## BAB III

### HASIL DAN PEMBAHASAN

#### 3.1 Dataset

Untuk penelitian ini, penulis menggunakan dataset distribusi Sari Roti yang didapatkan dari laporan distribusi mulai dari tanggal 1 hingga 9 Juni 2022.

SALESMAN ID	SALESMAN NIK	SALESMAN NAME	DROPPING NUMBER	DROPPING DATE	OUTLET BRAND	OUTLET CODE	OUTLET NAME	PRODUCT ID	PRODUCT CODE
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	19	FG101031
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	20	FG101032
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	137	FG101069
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	120	FG101074
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	166	FG101083
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	165	FG101084
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	126	FG203022
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	72	FG204001
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	153	FG206030
71299205	135441871299205	REZKI SAPUTRA	407134420522060115234830	2022-06-01	Sari Roti	1354390796	(HWK) ANDI KEC. PASAMAN	178	FG206032

Tabel 1 Dataset

SHORT NAME	QTY ESTIMASI	QTY DROPPING	DROPPING CBP	DISC. DROPPING	DROPPING RBP	QTY RETUR BAIK	RETUR BAIK CBP	RETUR BAIK RBP	DISC RETUR BAIK	QTY RETUR BS	RETUR BS CBP	RETUR BS RBP
RTSII	0	2	29000	2900	26100	0	0	0	0	0	0	0
RTGII	0	3	64500	6450	58050	0	0	0	0	0	0	0
SKK2	0	12	60000	6000	54000	0	0	0	0	0	0	0
SSM2	0	2	11000	1100	9900	0	0	0	0	0	0	0
RITS2	0	1	18000	1800	16200	0	0	0	0	0	0	0
RIKU II	0	2	39000	3900	35100	0	0	0	0	0	0	0
SRM2	0	1	6000	600	5400	0	0	0	0	0	0	0
DIC	0	1	5500	550	4950	0	0	0	0	0	0	0
RKS II	0	1	10500	1050	9450	0	0	0	0	0	0	0
TOC II 5S	0	6	105000	10500	94500	0	0	0	0	0	0	0

Tabel 2 Dataset

DISC RETUR BS	INVOICE DATE	INVOICE NUMBER	ONE TIME ARRIVAL	TYPE	OUTLET NAME
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI
0	2022-06-01	507134420522060115260592	2022-06-01 15:23:48	MR	ROTI

Tabel 3 Dataset

Penjelasan dari tiap kolom sebagai berikut :

#### 1. REPORT NAME

- Ini adalah nama laporan, yang menunjukkan jenis laporan yang dihasilkan atau konteksnya.pada kasus ini nama laporannya adalah “Detail Penjualan Mentah D9017”

#### 2. REPORT GENERATED TIME

- Waktu pembuatan laporan, menunjukkan kapan laporan tersebut dihasilkan atau diproses.
3. PLANT NAME
    - Nama pabrik atau fasilitas produksi yang terkait dengan data penjualan.
  4. ENTITY CODE
    - Kode entitas Distributor, mengacu pada kode unik yang mengidentifikasi entitas tertentu dalam sistem.
  5. ENTITY NAME
    - Nama entitas, merujuk pada nama atau deskripsi entitas distributor yang terkait dengan data penjualan.
  6. SALESMAN BRAND
    - Merek yang diwakili oleh salesman dalam penjualan tersebut, dalam hal ini Sariroti.
  7. SALESMAN ID
    - ID atau identifikasi unik untuk salesman yang terlibat dalam penjualan.
  8. SALESMAN NIK
    - Nomor induk karyawan (NIK) salesman yang terlibat dalam penjualan.
  9. SALESMAN NAME
    - Nama lengkap dari salesman yang terlibat dalam penjualan.
  10. DROPPING NUMBER
    - Nomor penurunan, merujuk pada nomor referensi untuk penurunan atau pengiriman produk ke outlet.
  11. DROPPING DATE
    - Tanggal penurunan atau pengiriman produk ke outlet.
  12. OUTLET BRAND
    - Merek outlet atau toko di mana produk dijual.
  13. OUTLET CODE
    - Kode unik yang mengidentifikasi outlet atau toko di mana produk dijual.
  14. OUTLET NAME
    - Nama outlet atau toko di mana produk dijual.
  15. PRODUCT ID

- ID atau identifikasi unik untuk produk yang terjual.

16. PRODUCT CODE

- Kode produk yang mungkin digunakan untuk mengidentifikasi produk tertentu.

17. PRODUCT NAME

- Nama lengkap produk yang terjual.

18. SHORT NAME

- Nama singkat atau singkatan untuk produk yang terjual.

19. QTY ESTIMASI

- Jumlah produk yang dipesan ke pabrik oleh distributor.

20. QTY DROPPING

- Jumlah produk yang dikirimkan atau diturunkan ke outlet.

21. DROPPING CBP

- Harga penurunan per unit berdasarkan CBP (Cost Before Price).

22. DISC. DROPPING

- Diskon yang diberikan pada produk yang diturunkan.

23. DROPPING RBP

- Harga penurunan per unit berdasarkan RBP (Recommended Base Price).

24. QTY RETUR BAIK

- Jumlah produk yang dikembalikan dalam kondisi baik.

25. RETUR BAIK CBP

- Harga per unit produk yang dikembalikan dalam kondisi baik berdasarkan CBP.

26. RETUR BAIK RBP

- Harga per unit produk yang dikembalikan dalam kondisi baik berdasarkan RBP.

27. DISC RETUR BAIK

- Diskon yang diberikan pada produk yang dikembalikan dalam kondisi baik.

28. QTY RETUR BS

- Jumlah produk yang dikembalikan dalam kondisi buruk atau cacat.

29. RETUR BS CBP

- Harga per unit produk yang dikembalikan dalam kondisi buruk atau cacat berdasarkan CBP.

30. RETUR BS RBP

- Harga per unit produk yang dikembalikan dalam kondisi buruk atau cacat berdasarkan RBP.

#### 31. DISC RETUR BS

- Diskon yang diberikan pada produk yang dikembalikan dalam kondisi buruk atau cacat.

#### 32. INVOICE DATE

- Tanggal penagihan atau tanggal faktur.

#### 33. INVOICE NUMBER

- Nomor faktur atau nomor referensi untuk transaksi penjualan.

#### 34. ONE TIME ARRIVAL

- Indikator apakah kunjungan outlet ini merupakan kunjungan satu kali atau reguler.

#### 35. TYPE OUTLET NAME

- Jenis atau kategori outlet atau toko di mana produk dijual.

## 3.2 Implementasi

### 3.2.1 Preprocessing Data

Pada tahap ini, dilakukan pengurangan kolom pada dataset untuk memastikan hanya fitur-fitur yang relevan digunakan dalam proses data mining menggunakan algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Langkah-langkah yang diambil dalam implementasi ini adalah sebagai berikut:

#### 1. Identifikasi Fitur Relevan:

- Dilakukan analisis awal terhadap dataset untuk mengidentifikasi fitur-fitur yang memiliki pengaruh signifikan terhadap proses clustering. Fitur-fitur yang dianggap tidak relevan atau memiliki korelasi rendah dengan tujuan analisis diabaikan.

#### 2. Pemilihan Dan Penyederhanaan Kolom yang Digunakan:

- Berdasarkan analisis yang telah dilakukan, kolom-kolom yang dipilih untuk digunakan dalam proses clustering adalah sebagai berikut:
  - **DROPPING DATE:** Tanggal penurunan barang.
  - **QTY DROPPING:** Jumlah barang yang diletakkan.
  - **QTY RETUR BS:** Jumlah barang yang dikembalikan.
  - **SHORT NAME:** Nama singkat produk.
  - **Name\_Code:** Kode nama produk.
  - **QTY SELLS:** Jumlah barang yang terjual. ( $QTY\ DROPPING - QTY\ RETUR\ BS$ )
  - **Day\_Code:** Kode hari. (Senin = 1)
  - **Date:** Tanggal transaksi.

DROPPING DATE	QTY DROPPING	QTY RETUR BS	SHORT NAME	Name_Code	QTY SELLS	Day_Code	Date
2022-06-01	2	0	RTSII	1	2	3	1
2022-06-01	3	0	RTGII	2	3	3	1
2022-06-01	12	0	SCK2	3	12	3	1
2022-06-01	2	0	SSM2	4	2	3	1
2022-06-01	1	0	RJTS2	5	1	3	1
2022-06-01	2	0	RJKU II	6	2	3	1
2022-06-01	1	0	SRM2	7	1	3	1
2022-06-01	1	0	DIC	8	1	3	1
2022-06-01	1	0	RKS II	9	1	3	1
2022-06-01	6	0	TOC II 5S	10	6	3	1
2022-06-01	2	0	TCS II 5S	11	2	3	1
2022-06-01	1	0	TST II 5S	12	1	3	1
2022-06-01	6	0	KKM	13	6	3	1
2022-06-01	3	0	SCCOR	14	3	3	1
2022-06-01	1	0	SCCK	15	1	3	1
2022-06-01	10	7	SCK2	3	3	3	1
2022-06-01	1	1	SSM2	4	0	3	1
2022-06-01	5	0	DIC	8	5	3	1

*Tabel 4 Dataset*



Name_Code	SHORT NAME	Name
1	RTSII	ROTI TAWAR SPESIAL II
2	RTGII	ROTI TAWAR GANDUM II
3	SCK2	SANDWICH COKLAT II
4	SSM2	SANDWICH PANDAN SARIKAYA MEDAN II
5	RJTS2	ROTI JUMBO TAWAR SPECIAL II
6	RJKU II	ROTI JUMBO TAWAR KUPAS II
7	SRM2	ROTI CREAM MOCCA II
8	DIC	DORAYAKI ISI COKLAT
9	RKS II	KASUR SUSU II
10	TOC II 5S	ROTI SOBEK COKLAT II 5S
11	TCS II 5S	ROTI SOBEK COKLAT SARIKAYA II 5S
12	TST II 5S	ROTI SOBEK COKLAT STRAWBERRY II 5S
13	KKM	ROTI KLASIK KASUR KRIM MESSES
14	SCCOR	STEAM CHEESE CAKE ORIGINAL
15	SCCK	STEAM CHEESE CAKE COKELAT
16	RKJ2	KASUR KEJU II
17	BKOR	BAUMKUCHEN ORIGINAL
18	BKCK	BAUMKUCHEN COKELAT
19	RKU2	ROTI TAWAR KUPAS II
20	RMS	ROTI TAWAR MILKY SOFT
21	SRC2	ROTI CREAM COKLAT II
22	SCC2	ROTI CREAM KEJU II
23	SCM	ROTI KRIM COKELAT MESES
24	TCC II 5S	ROTI SOBEK COKLAT KEJU II 5S
25	RSM2	ROTI SISIR MENTEGA II
26	SAP2	SANDWICH KRIM PEANUT II
27	SKJ2	SANDWICH KRIM KEJU II
28	DOS2	ROTI TAWAR DOUBLE SOFT II
29	RCC2	ROTI TAWAR CHOCO CHIP II
30	RTPDM2	ROTI TAWAR PANDAN MANIS II
31	ICK2	ROTI ISI COKLAT II
32	IKM	ROTI KLASIK ISI KRIM MESSES
33	SMG2	SANDWICH MARGARIN GULA II
34	TSC	SOBEK CHEESE & SPICY CHEESE
35	IKM3S	ISI KRIM MESSES FAMILY PACK 3 PCS

Table 5 kode roti

### 3. Transformasi dan Normalisasi Data:

- Data dalam kolom yang relevan kemudian ditransformasikan dan dinormalisasi untuk memastikan bahwa semua fitur berada dalam skala yang sama. Hal ini

penting untuk menghindari bias dalam algoritma DBSCAN, yang sensitif terhadap skala data.

Titik	Name_Code	QTY SELLS	Date
A	8	1	2
B	5	1	1
C	6	43	1
D	9	26	2
E	8	38	1
F	3	26	2
G	6	37	1
H	4	6	3
I	4	1	3
J	5	33	1
K	5	3	1

*Tabel 6 transformasi dan normalisasi data*

#### 4. Penggunaan Data yang Sudah Dibersihkan:

- Dataset yang sudah dibersihkan dari kolom-kolom yang tidak diperlukan kemudian digunakan sebagai input untuk algoritma DBSCAN. Proses ini memastikan bahwa hanya data yang signifikan yang mempengaruhi hasil clustering.

Proses pengurangan kolom ini terbukti penting dalam meningkatkan efisiensi dan akurasi dari hasil clustering menggunakan DBSCAN. Dengan menghilangkan fitur-fitur yang tidak relevan, algoritma dapat lebih fokus pada pola-pola penting dalam data, sehingga menghasilkan pembagian kluster yang lebih akurat dan bermakna.

### 3.3 Penerapan Algoritma DBSCAN dengan perhitungan manual.

Langkah-langkah yang dibutuhkan dalam segmentasi pelanggan menggunakan algoritma DBSCAN sebagai berikut. 1. Tentukan parameter Eps dan MinPts. Eps = 5 dan MinPts = 5

#### Iterasi 1

Pada iterasi I titik pusat awal ditentukan secara acak dengan demikian titik pusat adalah titik A dengan  $x_p=8$ ,  $y_p=1$ ,  $z_p=2$ . Berikut merupakan perhitungan jarak setiap titik dengan titik pusat.

$$\begin{aligned} AB &= \sqrt{(8-5)^2 + (1-1)^2 + (2-1)^2} \\ &= 3.16 \end{aligned}$$

$$\begin{aligned} AC &= \sqrt{(8-6)^2 + (1-43)^2 + (2-1)^2} \\ &= 42.05 \end{aligned}$$

$$\begin{aligned} AD &= \sqrt{(8-9)^2 + (1-26)^2 + (2-2)^2} \\ &= 25.10 \end{aligned}$$

$$\begin{aligned} AE &= \sqrt{(8-8)^2 + (1-38)^2 + (2-1)^2} \\ &= 37.01 \end{aligned}$$

$$\begin{aligned} AF &= \sqrt{(8-3)^2 + (1-26)^2 + (2-2)^2} \\ &= 25.49 \end{aligned}$$

$$\begin{aligned} AG &= \sqrt{(8-6)^2 + (1-37)^2 + (2-1)^2} \\ &= 36.06 \end{aligned}$$

$$\begin{aligned} AH &= \sqrt{(8-4)^2 + (1-6)^2 + (2-3)^2} \\ &= 6.48 \end{aligned}$$

$$\begin{aligned} AI &= \sqrt{(8-4)^2 + (1-1)^2 + (2-3)^2} \\ &= 4.12 \end{aligned}$$

$$\begin{aligned} AJ &= \sqrt{(8-5)^2 + (1-33)^2 + (2-1)^2} \\ &= 32.15 \end{aligned}$$

$$AK = \sqrt{(8-5)^2 + (1-3)^2 + (2-1)^2}$$

$$= 3.74$$

Setelah proses perhitungan maka dapat dilihat dalam bentuk tabel.

Titik	Jarak
AB	3.16227766
AC	42.05948169
AD	25.01999201
AE	37.01351105
AF	25.49509757
AG	36.06937759
AH	6.480740698
AI	4.123105626
AJ	32.15587038
AK	3.741657387

Tabel 7 hasil perhitungan iterasi 1

Density reachable diperoleh jika titik termasuk dalam radius dengan ketentuan nilai jarak kurang atau sama dengan Eps yang sudah ditetapkan. Berikut tabel titik terseleksi dengan ketentuan nilai Eps = 5.

Titik	Name_Code	QTY SELLS	Date	Jarak Ke Titik A
A	8	1	2	0
B	5	1	1	3.16227766
C	6	43	1	42.05948169
D	9	26	2	25.01999201
E	8	38	1	37.01351105
F	3	26	2	25.49509757
G	6	37	1	36.06937759
H	4	6	3	6.480740698
I	4	1	3	4.123105626
J	5	33	1	32.15587038
K	5	3	1	3.741657387

Tabel 8 titik terseleksi dengan eps = 5

Neighborhood Core Object tercapai lantaran total e-neighborhood telah melengkapi syarat karena total titik lebih dari sama dengan Min\_samples = 5. Titik selanjutnya dipilih titik yang tidak terkandung sebagai objek ataupun core, disini kami pilih titik G.

## Iterasi 2

Pada iterasi II titik pusat didapat dari iterasi I yaitu titik G dengan  $x_p=6$  ,  $y_p=37$  ,  $z_p=2$ . Berikut tabel hasil perhitungan jarak setiap titik dengan titik pusat pada iterasi II.

Titik	Jarak
GA	36.06937759
GB	36.01388621
GC	6
GD	11.44552314
GE	2.236067977
GF	11.44552314
GG	0
GH	31.12876483
GI	36.11094017
GJ	4.123105626
GK	34.0147027

Tabel 9 iterasi 2

Titik	Name_Code	QTY SELLS	Date	Jarak Ke Titik G
A	8	1	2	36.06937759
B	5	1	1	36.01388621
C	6	43	1	6
D	9	26	2	11.44552314
E	8	38	1	2.236067977
F	3	26	2	11.44552314
G	6	37	1	0
H	4	6	3	31.12876483
I	4	1	3	36.11094017
J	5	33	1	4.123105626
K	5	3	1	34.0147027

Tabel 10 jarak titik g

Dapat dilihat titik E dan J merupakan border point dari titik G karena belum pernah termasuk pada iterasi sebelumnya, titik yang dapat dipilih selanjutnya adalah titik I.

## Iterasi 3

Pada iterasi III titik pusat didapat dari iterasi II yaitu titik I dengan  $x_p=4$  ,  $y_p=1$  ,  $z_p=3$ . Berikut tabel hasil perhitungan jarak setiap titik dengan titik pusat pada iterasi III.

Titik	Jarak
IA	4.123105626
IB	2.236067977
IC	42.09513036
ID	25.51470164
IE	37.26929031
IF	25.03996805
IG	36.11094017
IH	5
II	0
IJ	32.07802986
IK	3

Tabel 11 iterasi 3

Titik	Name_Code	QTY SELLS	Date	Jarak Ke Titik I
A	8	1	2	4.123105626
B	5	1	1	2.236067977
C	6	43	1	42.09513036
D	9	26	2	25.51470164
E	8	38	1	37.26929031
F	3	26	2	25.03996805
G	6	37	1	36.11094017
H	4	6	3	5
I	4	1	3	0
J	5	33	1	32.07802986
K	5	3	1	3

Tabel 12 jarak ke titik i

Iterasi 1	Iterasi 2	Iterasi 3
AB	GE	IH
AI	GJ	
AK		

Tabel 13 cluster sementara dari setiap iterasi

Berdasarkan tabel dapat dilihat pembagian cluster sementara dari core point yang berbeda. Syarat untuk menjadi 1 cluster penuh adalah `min_samples = 5`. untuk menghitung seluruh data maka diperlukan otomatisasi menggunakan pemrograman python DBSCAN.

### 3.4 Proses Clustering dengan Algoritma DBSCAN

- a. Langkah awal adalah mengimport library-library yang diperlukan seperti pandas untuk membaca dan memanipulasi data, DBSCAN dari sklearn.cluster untuk melakukan clustering dengan algoritma DBSCAN, StandardScaler dari sklearn.preprocessing untuk melakukan normalisasi pada data, dan matplotlib.pyplot digunakan untuk memvisualisasikan data dalam bentuk plot.

```
import pandas as pd
import numpy as np
from sklearn.cluster import DBSCAN
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import
Axes3D

data =
pd.read_excel('sarirotidata2.xlsx')
df = pd.DataFrame(data)
data.head(20)
```

*Code 1 import library dan membaca dataset*

- b. Metode `info()` dipanggil pada `DataFrame` untuk mendapatkan informasi umum mengenai data, seperti jumlah baris, tipe data setiap kolom, dan jumlah nilai non-null di setiap kolom. Informasi ini membantu memahami struktur data yang sedang dikerjakan.

```
data.info()
```

*Code 2 info dataset*

- c. Kolom `Name_Code`, `QTY SELLS`, dan `Date` dipilih dari `DataFrame` dan diubah menjadi array numpy `X`. Objek `DBSCAN` kemudian dibuat dengan parameter `eps=5` dan `min_samples=5`. Parameter `eps` menentukan radius maksimum antara dua titik agar dianggap sebagai tetangga, sementara `min_samples` adalah jumlah minimum titik yang diperlukan untuk membentuk kluster. Algoritma `DBSCAN` dijalankan pada data `X` dan label kluster untuk setiap titik data disimpan dalam variabel `clusters`. Label ini kemudian ditambahkan sebagai kolom baru `'Cluster'` di `DataFrame` `df`.

```
X = df[['Name_Code', 'QTY SELLS', 'Date']].values
dbscan = DBSCAN(eps=5, min_samples=5)
clusters = dbscan.fit_predict(X)
df['Cluster'] = clusters
```

*Code 3 deklarasi dbscan*

- d. Plot 3D dibuat menggunakan `matplotlib`. Sebuah objek `figure` dengan ukuran 20x7 dibuat, kemudian sebuah `subplot 3D` ditambahkan ke dalamnya. Scatter plot 3D dibuat menggunakan data yang telah dikelompokkan, dengan `Name_Code` sebagai sumbu x, `QTY SELLS` sebagai sumbu y, dan `Date` sebagai sumbu z. Warna titik-titik dalam plot ditentukan oleh label kluster mereka, dengan menggunakan skema warna `'magma'`. Label untuk setiap sumbu dan judul untuk plot juga ditambahkan. Legenda plot dibuat untuk menunjukkan

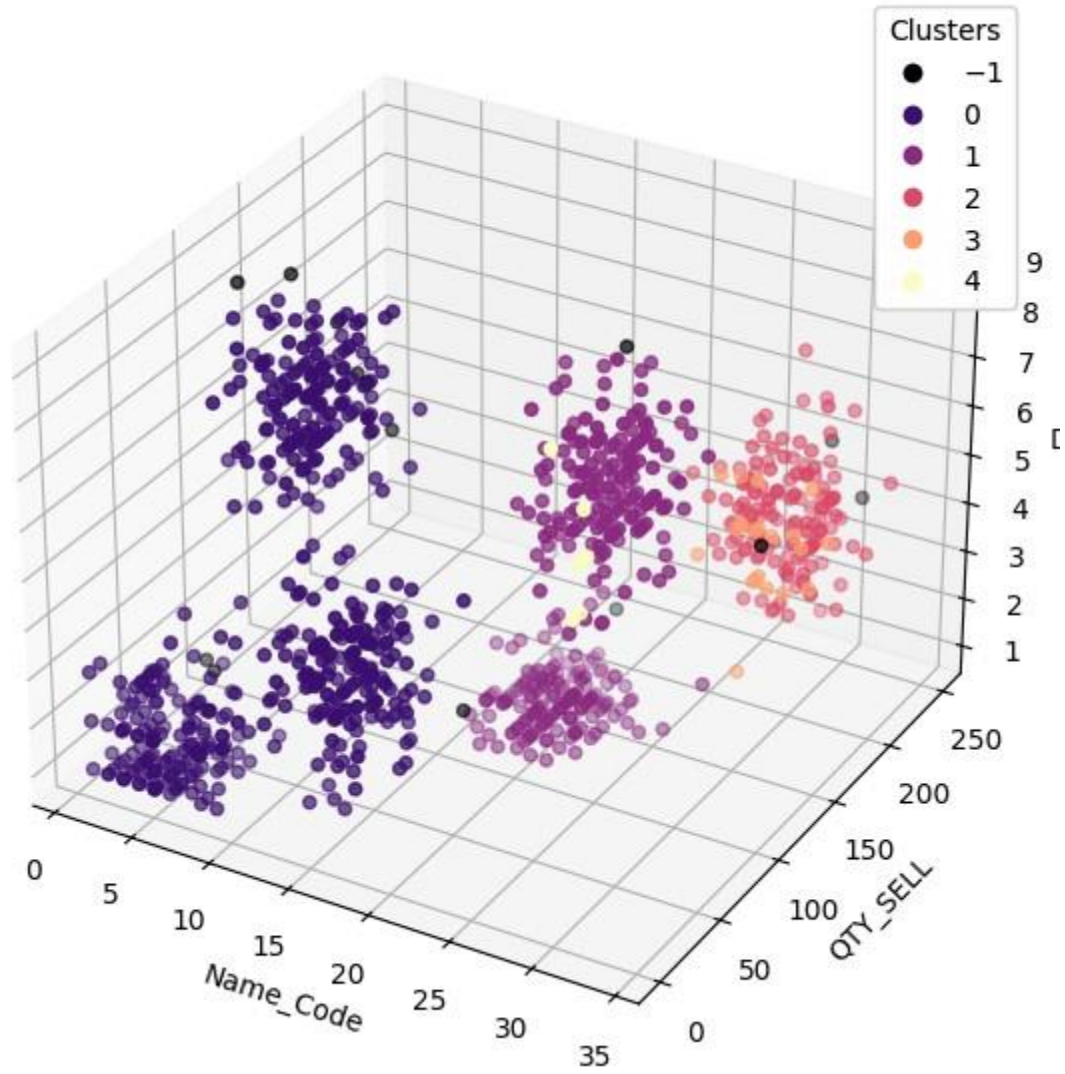


klaster yang berbeda dan ditambahkan ke plot menggunakan metode `add_artist`. Plot tersebut kemudian ditampilkan dengan `plt.show()`.

```
fig = plt.figure(figsize=(20, 7))
ax = fig.add_subplot(111, projection='3d')
scatter = ax.scatter(df['Name_Code'], df['QTY_SELLS'], df['Date'], c=df['Cluster'], cmap='magma')
ax.set_xlabel('Name_Code')
ax.set_ylabel('QTY_SELL')
ax.set_zlabel('Date')
ax.set_title('3D DBSCAN Clustering')
legend1 = ax.legend(*scatter.legend_elements(), title="Clusters")
ax.add_artist(legend1)
plt.show()
```

*Code 4 visualisasi cluster*

### 3D DBSCAN Clustering



Gambar 2 visualisasi cluster menggunakan 3d plot

- e. Defaultdict dari pustaka collections diimpor untuk membuat dictionary dengan nilai default berupa list. Klasterisasi dilakukan lagi menggunakan DBSCAN pada data X untuk memastikan konsistensi hasil. Dictionary cluster\_members dibuat untuk menyimpan indeks anggota tiap klaster. Dalam loop, dictionary diisi dengan menambahkan indeks tiap titik data ke dalam daftar yang sesuai dengan label klasternya. Untuk mencetak anggota dari tiap klaster, dilakukan pengecekan bahwa label klaster bukan noise (label != -1) dan bahwa jumlah anggota klaster memenuhi syarat minimum 5 anggota (len(members) >= 5). Ini memastikan bahwa hanya klaster valid yang dicetak, dengan menampilkan label klaster dan daftar indeks anggota dari setiap klaster.

```
from collections import defaultdict

dbscan = DBSCAN(eps=5, min_samples=5)
clusters = dbscan.fit_predict(X)

cluster_members = defaultdict(list)
for i, label in enumerate(clusters):
    cluster_members[label].append(i)

for label, members in
cluster_members.items():
    if label != -1 and len(members) >= 5:
        print(f'Cluster {label}:
{members}')
```

*Code 5 melihat list cluster*

```
cluster_counts = {label: len(members) for label, members in
cluster_members.items() if label != -1}

for label, count in cluster_counts.items():
    print(f'Cluster {label} memiliki {count} anggota')
```

*Code 6 jumlah anggota tiap cluster*

```
Cluster 0 memiliki 488 anggota
Cluster 1 memiliki 324 anggota
Cluster 2 memiliki 141 anggota
Cluster 3 memiliki 21 anggota
Cluster 4 memiliki 6 anggota
```

*Gambar 3 jumlah anggota tiap cluster*

Gambar 3D menunjukkan hasil klasterisasi penjualan roti berdasarkan kode roti (**Name\_Code**), jumlah roti terjual (**QTY SELLS**), dan tanggal penjualan (**Date**). Berikut adalah rincian tiap klaster:

- Cluster 0** (488 anggota): Roti dengan penjualan tinggi dan konsisten. Produk-produk ini sangat populer dan perlu pengelolaan stok yang baik untuk memenuhi permintaan pasar.
- Cluster 1** (324 anggota): Roti dengan penjualan signifikan namun bervariasi. Penjualan mungkin musiman atau berfluktuasi. Penting untuk menyesuaikan strategi pemasaran dengan tren penjualan.
- Cluster 2** (141 anggota): Roti dengan penjualan stabil tapi lebih rendah. Produk-produk ini memiliki pasar spesifik. Evaluasi strategi pemasaran dapat meningkatkan penjualan.
- Cluster 3** (21 anggota): Roti dengan penjualan rendah dan tidak reguler. Produk-produk ini perlu perhatian khusus dan evaluasi mendalam untuk menentukan langkah selanjutnya.
- Cluster 4** (6 anggota): Roti dengan penjualan sangat rendah. Perlu analisis untuk memahami alasannya dan mempertimbangkan apakah produk ini masih layak dipertahankan.

- f. **Noise (-1):** Data outlier yang tidak masuk klaster. Perlu analisis lebih lanjut untuk mengidentifikasi anomali atau tren baru.

## **BAB IV**

### **PENUTUP**

#### **4.1 Kesimpulan**

Dengan menggunakan algoritma DBSCAN, perusahaan distribusi Sari Roti dapat mengidentifikasi segmen pelanggan berdasarkan pola pembelian produk. Hasil clustering dapat digunakan untuk mengembangkan strategi pengelolaan stok dan distribusi yang lebih efektif, meningkatkan efisiensi operasional, dan meningkatkan kepuasan pelanggan. Implementasi data mining dalam analisis data penjualan juga memberikan wawasan yang berharga untuk pengambilan keputusan bisnis yang lebih baik.

Penelitian ini berhasil menerapkan algoritma DBSCAN untuk segmentasi pelanggan berdasarkan pola pembelian produk Sari Roti. Algoritma DBSCAN terbukti mampu menangani data dengan kepadatan yang bervariasi dan mengidentifikasi cluster dengan bentuk yang tidak beraturan. Hasil clustering menunjukkan pengelompokan pelanggan berdasarkan volume penjualan, jenis produk yang dibeli, dan waktu pembelian.

Dengan informasi ini, perusahaan distribusi Sari Roti dapat memperoleh pemahaman yang lebih baik tentang preferensi dan perilaku pembelian pelanggan di setiap gerai atau toko. Hal ini memungkinkan perusahaan untuk mengoptimalkan manajemen stok, merencanakan distribusi produk secara lebih efisien, dan menyesuaikan strategi pemasaran untuk memenuhi kebutuhan setiap segmen pelanggan.

#### **4.2 Saran**

Berhubungan dengan kesimpulan di atas, adapun saran yang dapat penulis berikan sebagai masukan adalah sebagai berikut:

1. Evaluasi berkala terhadap parameter `eps` dan `min_samples` yang digunakan dalam algoritma DBSCAN sangat penting untuk dilakukan. Hal ini bertujuan untuk memastikan bahwa parameter yang digunakan tetap optimal seiring dengan perubahan pola data penjualan yang mungkin terjadi dari waktu ke waktu. Parameter yang optimal akan menghasilkan segmentasi pelanggan yang lebih akurat dan sesuai dengan kondisi terkini.

2. Perusahaan perlu melakukan integrasi hasil segmentasi pelanggan dengan sistem manajemen distribusi yang dimiliki. Integrasi ini memungkinkan otomatisasi dalam pengaturan distribusi berdasarkan segmen pelanggan yang telah teridentifikasi. Proses ini akan meningkatkan efisiensi dan optimalisasi dalam proses distribusi secara keseluruhan.
3. Pemanfaatan visualisasi data seperti grafik atau peta juga dapat membantu pemahaman yang lebih baik tentang pola distribusi setiap segmen pelanggan, sehingga memfasilitasi pengambilan keputusan yang lebih baik dalam optimalisasi distribusi. Dalam upaya meningkatkan akurasi dan relevansi hasil segmentasi, perusahaan disarankan untuk melakukan analisis perbandingan antar segmen pelanggan yang dihasilkan. Dengan mengidentifikasi karakteristik unik setiap segmen dalam hal pola pembelian, perusahaan dapat menyesuaikan strategi distribusi dan pemasaran dengan lebih efektif.
4. Kolaborasi dengan tim lapangan seperti sales dan distributor juga diperlukan untuk menginterpretasikan hasil segmentasi secara lebih komprehensif. Mereka memiliki pengetahuan langsung tentang pelanggan dan kondisi lapangan, sehingga dapat memperkaya wawasan dan membantu penyesuaian strategi distribusi yang lebih tepat sasaran.

## DAFTAR PUSTAKA

- [1] Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.
- [2] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- [3] Tan, P. N., Steinbach, M., & Kumar, V. (2006). Introduction to data mining. Pearson Addison Wesley.
- [4] Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons.
- [5] Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques. Morgan kaufmann.
- [6] Kamus Besar Bahasa Indonesia (KBBI) Daring, "distribusi" dan "distributor", diakses pada 23 Mei 2024, dari <https://kbbi.kemdikbud.go.id/> .
- [7] Kotler, P., & Keller, K. L. (2016). Marketing Management (15th ed.). Pearson.
- [8] Aggarwal, C. C. (2015). Data mining: the textbook. Springer.
- [9] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651-666.
- [10] Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2), 165-193.
- [11] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, 96(34), 226-231.
- [12] Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 1-21.