

Penerapan Algoritma K-Means Clustering untuk Optimalisasi Persediaan Liquid Vape Berdasarkan Data Penjualan

Selfi Melisa[✉], Sarjon Defit, Rini Sovia

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia

Selfimelisa.22@gmail.com

Abstract

Liquid vape is a liquid in an electronic cigarette (vape) device that contains a mixture of Propylene Glycol (PG), Vegetable Glycerin (VG), flavorings, and contains nicotine. As the use of vapes increases as an alternative to conventional cigarettes, efficient stock management becomes a challenge for vape shops to be able to meet customer needs without experiencing excess or shortage of inventory. Good stock management in a retail business is very important to maintain a balance between demand and product availability. This research aims to optimize liquid vape supplies by analyzing sales patterns. This research method is K-Means Clustering which includes several stages, namely determining the number of clusters, determining the centroid point randomly, calculating the closest distance between data and the centroid using the Euclidean method, grouping data into each cluster, updating the centroid until it is stable, and evaluating the results. The data used in the research is liquid vape sales data from June to November 2024 with a total of 68 product samples. Data processing was carried out manually and testing used RapidMiner software to measure the level of accuracy of the clustering results. The research results show that the K-Means Clustering algorithm is successful in grouping products into three categories: very popular, best selling, and not very popular. 51 products are in the low-selling category, 13 products are in the best-selling category, and 4 products are in the very best-selling category, with a Davies Bouldin value of 0.374%. The application of K-Means Clustering is effective in grouping products according to demand, helps determine the ideal stock amount, reduces the risk of product excesses or shortages, and increases operational efficiency.

Keywords: Liquid Vape, Data Mining, K-Means Clustering, Stock Management, Sales Pattern Analysis.

Abstrak

Vegetable Glycerin (VG), perasa, dan terkandung nikotin. Seiring meningkatnya penggunaan vape sebagai alternatif rokok konvensional, pengelolaan stok yang efisien menjadi tantangan bagi toko vape agar dapat memenuhi kebutuhan pelanggan tanpa mengalami kelebihan atau kekurangan persediaan. Manajemen stok yang baik dalam bisnis ritel sangat penting untuk menjaga keseimbangan antara permintaan dan ketersediaan produk. Penelitian ini bertujuan untuk mengoptimalkan persediaan liquid vape dengan menganalisis pola penjualan. Metode penelitian ini yaitu K-Means Clustering dengan mencakup beberapa tahapan yaitu menentukan jumlah klaster, menentukan titik centroid secara acak, menghitung jarak terdekat antar data dengan centroid menggunakan metode Euclidean, mengelompokkan data ke masing-masing cluster, pembaruan centroid hingga stabil, dan evaluasi hasil. Data yang digunakan dalam penelitian yaitu data penjualan liquid vape dari bulan Juni hingga November 2024 dengan total 68 sampel produk. Pengolahan data dilakukan dengan cara manual dan pengujiannya menggunakan software RapidMiner dalam mengukur tingkat akurasi hasil clustering. Hasil penelitian menunjukkan bahwa algoritma K-Means Clustering berhasil mengelompokkan produk ke dalam tiga kategori: sangat laris, laris, dan kurang laris. 51 produk masuk ke dalam kategori kurang laris, 13 produk masuk ke dalam kategori laris, dan 4 produk masuk ke dalam kategori sangat laris, dengan nilai Davies Bouldin sebesar 0.374%. Penerapan K-Means Clustering efektif dalam mengelompokkan produk sesuai permintaan, membantu menentukan jumlah stok yang ideal, mengurangi resiko kelebihan maupun kekurangan produk, serta meningkatkan efisiensi operasional.

Kata kunci: Liquid Vape, Data Mining, K-Means Clustering, Manajemen Stok, Analisa Pola Penjualan

KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



1. Pendahuluan

Pertumbuhan dunia usaha di era globalisasi dan kemajuan pesat teknologi informasi memberikan dampak yang besar baik pada sektor industri maupun jasa [1]. Penggunaan teknologi Big Data dalam bisnis ritel telah menjadi semakin penting dalam menghadapi persaingan yang ketat di pasar global saat ini. Teknologi Big Data memungkinkan perusahaan ritel untuk mengumpulkan, mengelola, dan menganalisis

data besar dengan cepat dan efisien, yang gilirannya dapat mempengaruhi keputusan konsumen [2]. Dalam konteks perkembangan teknologi yang cepat, manajemen persediaan barang menjadi semakin penting, terutama dalam industri, dimana ketersediaan dan kelengkapan barang memiliki dampak langsung pada kinerja perusahaan. Tantangan yang muncul adalah bagaimana mengelola stok barang dengan efisien agar selalu tersedia sesuai dengan permintaan,

yang memerlukan pendekatan yang cermat dan terstruktur [3].

Manajemen persediaan barang yang efisien dan efektif memerlukan implementasi sistem informasi yang andal. Optimalisasi pengelolaan inventaris, perusahaan perlu memiliki kemampuan dalam mengelola jumlah stok barang sesuai dengan kebutuhan [4]. *Knowledge Discovery in Databases* (KDD) merupakan metode teknis yang handal dan berguna untuk mencari serta mengidentifikasi pola didalam sebuah data [5]. KDD melibatkan serangkaian langkah dan teknik untuk mengidentifikasi pola, tren, dan informasi penting lainnya dari data besar dengan tujuan mendukung pengambilan keputusan yang lebih baik [6]. Proses KDD terdapat 5 tahapan yaitu tahap seleksi data, *pre-processing*, transformasi, proses mining dan evaluasi [5].

Clustering merupakan salah satu metode dalam data mining yang digunakan untuk mengelompokkan ke dalam kelompok atau klaster berdasarkan karakteristik atau atribut tertentu untuk mengidentifikasi pola data [7]. Salah satu teknik yang sering digunakan data mining dalam pengelompokan data adalah *K-Means Clustering*. Algoritma *K-Means* merupakan metode *clustering* dengan *non hirarki* yang mempartisi data-data ke dalam *cluster* dimana data-data dengan karakteristik yang sama akan dikelompokkan pada satu *cluster*, dan data-data dengan karakteristik yang berbeda akan dikelompokkan pada *cluster* lainnya [8]. Penelitian sebelumnya dilakukan oleh Astuti & Muqorobin tentang *K-Means Clustering* yang digunakan untuk pengelompokan obat di Puskesmas Mertoyudan I Magelang. Penelitian ini menghasilkan aplikasi pengelompokan data obat yang mampu mengelompokkan tiga *cluster* yaitu pada *cluster* obat ringan terdapat 5 obat, pada *cluster* obat sedang terdapat 11 obat dan pada *cluster* obat berat terdapat 4 obat. Pada uji validitas diperoleh hasil yang cocok antara perhitungan manual algoritma *k-means cluster* dengan hasil program aplikasi dan terakhir pada uji model Silhouette Coefficient (SC) maka diperoleh hasil rata-rata yang baik yakni 0,97 hampir mendekati angka 1 sehingga sistem ini dinyatakan *strang structure* [9].

Penelitian selanjutnya dilakukan oleh Djaka Permana dkk tentang *K-Means Clustering* digunakan untuk klasterisasi data jamaah umrah pada Tanurmutmainnah Tour. Penelitian ini menggunakan perangkat lunak RapidMiner 9.9 untuk menghitung data, dan hasilnya menunjukkan bahwa 38% kelompok sangat diminati dengan paket kamar Quad, 34% kelompok diminati dengan paket kamar Triple, dan 28% kelompok kurang diminati untuk paket kamar Double. Temuan ini memberikan informasi terbaru kepada manajer di Tanurmutmainnah yang dapat membantu mereka menerapkan strategi pemasaran [10].

Penelitian berikutnya dilakukan oleh Fikria & Sriani 2024 tentang *K-Means Clustering* digunakan dalam

pengelompokan penjualan sembako di PT. Sinarmas Distribusi Nusantara. Dengan mengelompokkan data penjualan PT. Sinarmas Distribusi Nusantara, perusahaan dapat mengidentifikasi produk-produk yang paling diminati oleh setiap *cluster* konsumen. Proses pengelompokan data memperoleh hasil bahwa dari keseluruhan produk yang diteliti, 5 produk termasuk pada kategori “Laris”, 2 produk pada kategori “Kurang Laris”, dan 51 produk pada kategori “Tidak Laris”. Hasil dari proses pengelompokan ini memperoleh nilai *davies-bouldin index* sebesar 0.8911 [11].

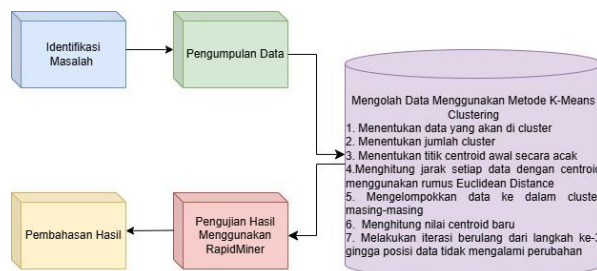
Penelitian yang saman juga dilakukan oleh Triyandana dkk 2022, tentang algoritma *K-Means* digunakan untuk mengelompokkan menu makanan dan minuman berdasarkan tingkat penjualan dari DPom Coffee. Berdasarkan hasil pengolahan dataset penjualan DPom Coffee menggunakan *tools RapidMiner*, menghasilkan 3 *cluster* dengan nilai DBI sebesar -0.457, yaitu *cluster* 1 menu dengan tingkat penjualan rendah sebanyak 8, *Cluster* 2 dengan menu memiliki tingkat penjualan sedang sebanyak 40, dan untuk *cluster* 3 sebanyak 7 menu dengan tingkat penjualan tinggi [12].

Penelitian yang senada juga dilakukan oleh Aranski dkk 2024, tentang *K-Means* yang digunakan untuk mengelompokkan data penerima bantuan subsidi rumah. Jumlah dataset sebanyak 170 dibagi kedalam 2 *cluster* yaitu layak dan tidak layak mendapatkan bantuan subsidi rumah. Hasil pengujian dapat disimpulkan bahwa *cluster* 0 atau layak mendapatkan bantuan subsidi rumah adalah sebanyak 91 kepala keluarga, dan *cluster* 1 atau tidak layak ada sebanyak 79 kepala keluarga tidak layak mendapatkan bantuan [13].

Berdasarkan penelitian terdahulu diatas menunjukkan bahwa *K-Means Clustering* telah banyak digunakan untuk pengelompokan data dalam berbagai industri. Penggunaan metode *Clustering* dalam industri seperti penjualan vape masih sangat terbatas sehingga pengklasteran data penjualan vape diperlukan. Hal ini seiring dengan pihak industri vape memerlukan untuk mengetahui pola konsumsi yang dinamis, sehingga sistem manajemen stok berbasis data sangat diperlukan untuk meningkatkan efisiensi bisnis.

2. Metodologi Penelitian

Metodologi dalam penelitian ini menggunakan pendekatan kualitatif dengan menerapkan konsep data mining dengan algoritma *K-Means* yang digunakan untuk optimalisasi persediaan liquid vape berdasarkan data penjualan. Metodologi penelitian ini menjelaskan rancangan kegiatan penelitian secara sistematis dengan langkah-langkah yang telah ditentukan. Adapun gambaran kerangka kerja penelitian dapat dilihat pada Gambar 1.



Gambar 1 Kerangka Kerja Penelitian

Gambar 1 merupakan alur tahap penelitian dalam proses optimalisasi persediaan liquid vape. Proses kerja penelitian ini dimulai dengan melakukan identifikasi masalah, pengumpulan data, proses analisis K-Means Clustering dengan menggunakan alat bantu RapidMiner. Hasil proses analisis akan mampu menyajikan pola terhadap penjualan liquid vape yang dijadikan bentuk optimalisasi dalam manajemen persediaan.

2.1 Data Mining

Data mining merupakan metode untuk menemukan informasi tersembunyi dalam *database* dan bagian dari proses *Knowledge Discovery in Databases* (KDD) untuk menemukan informasi dan pola yang berguna dalam data [14]. Data mining sendiri merupakan proses menganalisa data agar dapat menemukan suatu pola yang ada di dalam sekumpulan data. Data mining juga memiliki kemampuan untuk menganalisa data yang sangat besar menjadi informasi berupa pola yang memiliki arti sehingga dapat digunakan oleh pengambil keputusan [8]. Data mining memiliki kelompok penyelesaian berdasarkan permasalahan masing-masing diantaranya [7]:

1. Klasifikasi
Klasifikasi merupakan proses pengklasifikasian dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas yang telah ditentukan sebelumnya.
2. Clustering
Clustering merupakan metode dalam data mining yang digunakan untuk mengelompokkan data ke dalam kelompok atau klaster berdasarkan karakteristik atau atribut tertentu.
3. Asosiasi
Asosiasi merupakan konsep dalam data mining yang berkaitan dengan keterkaitan atau hubungan antara item atau atribut dalam suatu dataset.
4. Estimasi
Estimasi merupakan proses perhitungan atau perkiraan nilai atau jumlah yang tidak diketahui berdasarkan informasi atau data yang ada.
5. Prediksi
Prediksi merupakan proses atau hasil kerja dari melakukan estimasi atau perkiraan nilai atau kejadian yang akan terjadi di masa depan berdasarkan informasi atau data yang tersedia saat ini.

2.2 K-Means Clustering

K-Means Clustering merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melaksanakan pengelompokan atau *clustering* suatu data. Algoritma K-Means Clustering merupakan metode *clustering* dengan *non hirarki* yang mempartisi data-data ke dalam *cluster* dimana data-data dengan karakteristik yang sama dikelompokkan pada satu *cluster*, dan data-data dengan karakteristik yang berbeda akan dikelompokkan pada *cluster* lainnya [8]. Tahapan dalam K-Means Clustering yaitu sebagai berikut [15]:

1. Menentukan jumlah klaster
Langkah pertama dalam *K-Means* adalah menentukan jumlah klaster atau kelompok yang akan dibuat. Jumlah klaster dipilih berdasarkan pada tujuan dari *clustering* dan karakteristik data yang akan dianalisis.
 2. Memilih *centroid* awal
Setelah jumlah klaster ditentukan, langkah selanjutnya adalah memilih titik *centroid* awal secara acak dari seluruh data. Titik *centroid* inilah yang akan menjadi pusat dari setiap kelompok yang akan dibuat.
 3. Menghitung jarak
Setelah titik klaster ditentukan, langkah selanjutnya adalah menghitung jarak antara observasi dengan titik *centroid*. Jarak ini biasanya dihitung dengan menggunakan metrik *Euclidean* yang ditunjukkan pada persamaan.

$$Dist = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$
- Dimana:
- Dist* = adalah Eucclidean Distance
 - i* = adalah banyaknya objek
 - (x,y) = merupakan koordinat objek dan
 - (s,t) = merupakan koordinat *centroid*
4. Mengelompokkan data
Setelah jarak dihitung, langkah selanjutnya adalah mengelompokkan setiap data ke dalam kelompok yang paling dekat dengan titik *centroid*. Tiap data hanya bisa masuk ke dalam satu kelompok saja.
 5. Menghitung titik *centroid* baru
Setelah semua data dikategorikan, langkah selanjutnya adalah menghitung ulang titik *centroid* baru masing-masing kelompok berdasarkan rata-rata dari setiap data pada kelompok tersebut.

Setelah titik *centroid* baru dihitung, langkah selanjutnya adalah mengulangi proses langkah 3-5 dengan menggunakan titik *centroid* yang baru sebagai pusat dari kelompok-kelompok tersebut. Prosedur ini diulang terus menerus hingga titik *centroid* tidak berubah atau jumlah iterasi yang telah ditentukan telah tercapai.

3. Hasil dan Pembahasan

Pada tahap hasil dan pembahasan, data-data yang diperlukan untuk penelitian yaitu data penjualan/transaksi Vaporate Store. Metode K-Means Clustering ini mengolah data penjualan liquid dengan melakukan proses *clusterisasi*, sehingga hasil pengelompokan data dapat menjadi acuan pihak Vaporate Store. Analisa data dan pengolahan angka

diolah menggunakan *software* pendukung yaitu menggunakan *Microsoft Office Excel*. Setelah diolah dengan *Microsoft Office Excel* maka selanjutnya akan disimulasikan dengan menggunakan *software RapidMiner* untuk membandingkan hasil yang diolah dengan *Microsoft Office Excel*. Kemudian berdasarkan kerangka kerja penelitian yang sudah ditentukan sebelumnya, sehingga dibentuk bagan alir analisa dan perancangan yang terdapat pada Gambar 2.



Gambar 2. Bagan alir Analisa dan Perancangan

Gambar 2 diatas menunjukkan bagan alir analisa dan perancangan dalam penelitian ini. Proses dilakukan untuk mengoptimalkan pengelolaan stok atau persediaan liquid vape berdasarkan data penjualan/pola penjualan. Berikut penjelasan dari setiap tahapan dalam bagan alir:

3.1 Data Penjualan Liquid Bulan Juni hingga November 2024

Data yang akan digunakan dalam penelitian ini adalah data penjualan liquid bulan Juni hingga November 2024, dimana didapatkan data sebanyak 68 data yang dapat dilihat pada Tabel 1.

Tabel 1. Data Total Penjualan Bulan Juni hingga November 2024

No	Nama Produk	Juni	Juli	Agustus	September	Oktober	November	Total Penjualan	Rata-Rata Penjualan
1	Tetra Crackz 60 ml	4	2	8	10	5	8	37	6
2	Happi Crunch	2	7	0	3	6	1	19	3
3	Creamsie Whimsie	5	9	3	2	4	8	31	5
4	Lunar Ice Cream Series	18	7	11	22	8	14	80	13
5	Creaminal	10	0	9	3	2	2	26	4
...
64	Lunar Ice Cream Series Salt	15	4	3	7	19	14	62	10
65	Ala Carte Salt	1	0	0	5	3	0	9	2
66	Ice Pedia	2	1	0	1	0	53	57	10
67	Una Salt	4	6	12	10	10	5	47	8
68	Samose	6	1	0	2	0	0	9	2

Kemudian dari data diatas data yang akan diolah menggunakan *K-Means Clustering* adalah total penjualan dan rata-rata penjualan, yang mempresentasikan jumlah keseluruhan penjualan selama 6 bulan dari Juni hingga November 2024. Data total penjualan dapat dilihat pada Tabel 2.

64	Lunar Ice Cream Series Salt	62	10
65	Ala Carte Salt	9	2
66	Ice Pedia	57	10
67	Una Salt	47	8
68	Samose	9	2

Tabel 2. Data Total Penjualan Liquid Bulan Juni hingga November 2024

No	Nama Produk	Total Penjualan	Rata-Rata Penjualan
1	Tetra Crackz 60 ml	37	6
2	Happi Crunch	19	3
3	Creamsie Whimsie	31	5
4	Lunar Ice Cream Series	80	13
5	Creaminal	26	4
...

3.2 Analisa Data dengan Metode K-Means Clustering

Analisa proses algoritma *K-Means Clustering* dilakukan setelah dilakukan semua tahapan *preprocessing data* selesai. Pada analisa proses ini dilakukan pengolahan data penjualan liquid dengan menggunakan Algoritma *K-Means Clustering* sampai didapatkan hasil klaster untuk setiap liquid. Tahapan algoritma *K-Means Clustering* sebagai berikut:

1. Masukkan Data Yang Akan di Cluster

Data yang akan dimasukkan dalam penelitian ini adalah data total penjualan liquid vape dari bulan Juni hingga November 2024. Data tersebut mencakup berbagai produk dengan jumlah penjualan berbeda-beda selama periode tersebut. Informasi penjualan ini dapat dilihat secara rinci pada Tabel 2. Data ini akan dianalisis lebih lanjut untuk mengelompokkan produk berdasarkan data penjualan menggunakan K-Means Clustering.

2. Menentukan Jumlah Cluster

Menentukan jumlah cluster dalam K-Means Clustering adalah langkah penting yang mempengaruhi akurasi dan interpretasi hasil pengelompokan data. Pemilihan jumlah cluster yang tepat memastikan data dikelompokkan secara optimal berdasarkan pola yang ada, sehingga dapat digunakan dalam analisis lebih lanjut. Pada penelitian ini klasterisasi data menjadi 3 cluster, yaitu C0 (Kurang Laris), C1 (Laris), C2 (Sangat Laris).

3. Menentukan Titik Centroid Awal Secara Acak

Penentuan nilai *centroid* awal ini dapat diambil secara acak dari 68 data sampel yang digunakan dalam penelitian. *Centroid* awal ini merupakan titik pusat *cluster* pertama. *Centroid* untuk C0 diambil dari data sampel data ke 8, C1 diambil dari data sampel data ke 15 dan C2 diambil dari data sampel data ke 44 yaitu. Berikut merupakan bentuk *centroid* awal yang telah dituangkan kedalam bentuk Tabel 3.

Tabel 3. Centroid Awal

Centroid Awal	X	Y
C0	10	2
C1	20	3
C2	64	11

Pada Tabel 3 menunjukkan centroid awal yang digunakan dalam proses K-Means Clustering. Dalam tabel ini terdapat dua parameter yaitu x dan y, x merupakan total penjualan dan y merupakan rata-rata penjualan. Dengan nilai C0, C1, dan C2 diambil secara acak.

4. Menghitung Jarak Setiap Data dengan Centroid Menggunakan Rumus Euclidean Distance

Setelah dilakukan jumlah cluster dan nilai centroid awal pada proses algoritma K-Means Clustering, maka selanjutnya dilakukan perhitungan menggunakan rumus Euclidean Distance untuk menghitung jarak data dengan centroid. Perhitungannya menggunakan rumus (1) pada bagian metodologi penelitian. Proses perhitungan diulangi sampai jarak dari data ke-68 data tersebut terhadap centroid. Proses clustering pada data total penjualan liquid bulan juni hingga November

2024 menghasilkan 11 iterasi. Untuk hasil dari perhitungan pada iterasi 1 dapat dilihat pada Tabel 4.

Tabel 4. Hasil Kalsterisasi Nilai Cluster Pada Iterasi 1

No	Nama Produk	C0	C1	C2	Cluster
1	Tetra Crackz 60 ml	27,2946	17,2626	27,4590	C1
2	Happi Crunch	9,0553	1	45,7055	C1
3	Creamsie Whimsie	21,2132	11,1803	33,5410	C1
4	Lunar Ice Cream Series	70,8590	60,8276	16,1245	C2
5	Creamina 1	16,1245	6,0827	38,6393	C1
...
64	Lunar Ice Cream Series Salt	52,6117	42,5793	2,2360	C2
65	Ala Carte Salt	1	11,0453	55,7314	C0
66	Una Salt	37,4833	27,4590	17,2626	C2
67	Ice Pedia	47,6759	37,6563	7,0710	C2
68	Samose	1	11,0453	55,7314	C0

Tabel 4 menunjukkan hasil proses iterasi pertama dari proses K-Means Clustering, di mana setiap objek data dikelompokkan ke dalam cluster berdasarkan jarak terdekat dengan centroid. Pada iterasi pertama, centroid awal digunakan sebagai acuan untuk menentukan klasterisasi awal data penjualan. Setelah setiap objek data dikelompokkan, centroid akan diperbarui berdasarkan rata-rata posisi anggota cluster untuk iterasi berikutnya hingga mencapai kondisi yang stabil.

5. Mengelompokkan Data ke Dalam Cluster Masing-Masing

Tahap ini menyajikan hasil pengelompokan data ke dalam cluster pada iterasi 1. Proses pengelompokan data dilakukan dengan menentukan keanggotaan cluster berdasarkan jarak antara setiap data dengan centroid. Data-data yang memiliki jarak paling dekat dengan centroid yang sama akan dikelompokkan ke dalam cluster yang sama. Hasil pengelompokan iterasi dapat dilihat pada Tabel 5.

Tabel 5. Pengelompokan Hasil Iterasi 1

Cluster	Anggota	Jumlah
C0	8, 13, 16, 29, 31, 47, 50, 53, 61, 65, 68	11
C1	1, 2, 3, 5, 6, 10, 11, 12, 15, 18, 19, 20, 23, 24, 25, 26, 27, 33, 34, 35, 37, 38, 39, 40, 41, 42, 44, 46, 48, 49, 51, 52, 54, 55, 57, 58, 61, 62, 63	39
C2	4, 7, 9, 14, 17, 21, 22, 28, 30, 32, 36, 43, 45, 56, 59, 64, 66, 67	18

Tabel 5 menunjukkan hasil pengelompokkan pada iterasi 1. Dengan hasil yang didapatkan yaitu cluster C0 (Kurang Laris) sebanyak 11 anggota, C1 (Laris) sebanyak 39 anggota, dan C2 (Sangat Laris) sebanyak 18 anggota. Proses iterasi berlanjut hingga pengelompokkan hasil iterasi mencapai kondisi konvergensi, dimana posisi centroid dan keanggotaan cluster tidak lagi mengalami perubahan signifikan.

6. Menghitung Nilai Centroid Baru

Menghitung nilai centroid baru didapatkan dari menghitung rata-rata dari setiap data yang masuk dalam cluster yang di dapatkan dari iterasi 1. Berdasarkan hasil perhitungan rata-rata tersebut, diperoleh nilai centroid baru yang akan digunakan untuk iterasi berikutnya dalam proses K-Means Clustering. Nilai centroid yang diperbarui ini dapat dilihat pada Tabel 6 sebagai acuan untuk menentukan kembali pengelompokkan data pada iterasi selanjutnya.

Tabel 6. Centroid Baru Iterasi 2

Centroid Iterasi 2	X	Y
C0	10,4545	1,9090
C1	24,4871	4,1794
C2	85,7777	14,3888

Setelah dilakukan proses perhitungan nilai centroid baru yang didapatkan dari rata-rata setiap data yang masuk ke dalam cluster yang didapatkan pada iterasi 1. Didapatkan hasil nilai centroid untuk iterasi 2 dengan nilai x (total penjualan) sebesar 10,4545 dan y (rata-rata penjualan) sebesar 1.9090 untuk C0 (kurang laris). Kemudian nilai x (total penjualan) sebesar 24,4871 dan y (rata-rata penjualan) sebesar 4,1794 untuk C1 (laris). Terakhir nilai x (total penjualan) sebesar 85,7777 dan y (rata-rata penjualan) sebesar 14,3888 untuk C2 (sangat laris).

7. Melakukan Iterasi Berulang Kali Hingga Posisi Data Tidak Berubah

Setelah diperoleh centroid baru maka dilakukan iterasi 2 untuk mengetahui perhitungan jarak dari setiap data ke centroid baru dengan rumus *Euclidean Distance*. Perhitungan pada iterasi 2 sama langkahnya dengan perhitungan pada iterasi 1, tetapi pada iterasi berikutnya centroid yang digunakan adalah centroid yang terbaru. Setelah proses perhitungan diulangi sampai jarak dari data 68 data tersebut terhadap centroid. hasil dari masing-masing perhitungan untuk iterasi 2 ini akan diletakkan sesuai dengan urutan klasternya. Kemudian perhitungan ini berhenti di iterasi 11 karena tidak ada posisi data yang berubah. Hasil dari iterasi 11 dapat dilihat pada Tabel 7.

Tabel 7. Hasil Iterasi 11

No	Nama Produk	C0	C1	C2	Cluster
1	Tetra Crackz 60 ml	15,3431	30,2252	122,9707	C0

2	Happi Crunch	2,92013	48,4730	141,2190	C0
3	Creamsie Whimsie	9,2619	36,3078	129,0535	C0
4	Lunar Ice Cream Series	58,9080	13,3485	79,4059	C1
5	Creamina 1	4,1842	41,4054	134,1503	C0
...
64	Lunar Ice Cream Series Salt	40,6604	4,9254	97,6540	C1
65	Ala Carte Salt	12,9416	58,5020	151,2475	C0
66	Una Salt	35,7282	9,8464	102,5856	C1
67	Ice Pedia	25,5335	20,0314	112,7777	C1
68	Samose	12,9416	58,5020	151,2475	C0

Penjelasan tabel diatas menunjukkan bahwa hasil iterasi berhenti pada iterasi ke 11. Posisi cluster terakhir tidak berubah lagi dan ditemukan pada iterasi ke 11. Hasil dari perhitungan di kelompokkan ke dalam cluster nya masing-masing. Adapun hasil pengelompokkan dapat dilihat pada Tabel 8.

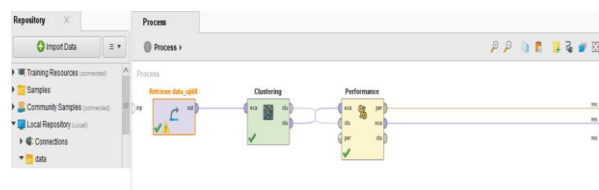
Tabel 8. Hasil Akhir Klasterisasi Iterasi 11

C0	1, 2, 3, 5, 6, 8, 10,11, 12,13, 15, 16, 19, 20, 21, 24, 25, 26, 27, 28, 30, 32, 34, 35, 36, 38, 39, 40, 41, 42, 43, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 63, 64, 66, 69	51
C1	4, 7, 9, 14, 18, 22, 23, 29, 33, 37, 44, 46, 65, 67, 68, 70	13
C2	17, 31, 57, 60	4

Tabel 8 menunjukkan hasil akhir klasterisasi yang berhenti pada iterasi 11 setelah tidak ada perubahan anggota dalam setiap cluster. Hasil yang diperoleh menunjukkan bahwa cluster C0 (Kurang Laris) sebanyak 51 anggota, C1 (Laris) sebanyak 13 anggota, dan C2 (Sangat Laris) sebanyak 4 anggota. Proses klasterisasi berhenti karena nilai centroid sudah stabil, sehingga tidak ada lagi perpindahan data antar cluster dalam iterasi berikutnya.

3.3 Pengujian dengan software RapidMiner

Dalam pengujian algoritma K-Means Clustering menggunakan software *RapidMiner*, dataset pada Tabel 2 dijadikan sebagai data yang akan diuji yang tersedia dalam format file Excel. Data tersebut kemudian akan diimpor ke dalam RapidMiner untuk memulai proses pengujian. Langkah-langkah pengujian yang dilakukan dapat divisualisasikan pada Gambar 3.



Gambar 3. Tampilan Input Dataset dan Pengujian Menggunakan Performance

Pada gambar diatas merupakan tampilan proses dalam software RapidMiner dalam analisis data dengan K-Means Clustering. Langkah pertama kita memasukkan dataset yang akan di proses dalam analisis klasterisasi, kemudian diproses menggunakan metode K-Means untuk mengelompokkan data berdasarkan kesamaan tertentu. Setelah proses klasterisasi selesai, hasilnya dievaluasi menggunakan *Performance Evaluation* untuk melihat kualitas klasterisasi dengan metrik tertentu seperti *Davies-Bouldin Index* atau *Sum of Squared Errors (SEE)*. Selanjutnya hasil dari proses K-Means Clustering dapat dilihat pada Gambar 4.

```

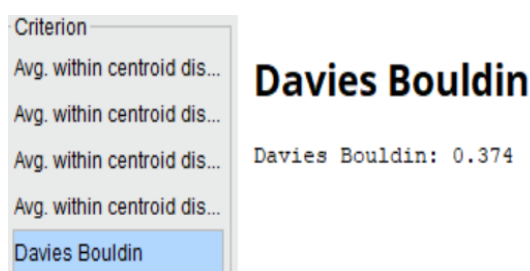
Cluster Model

Cluster 0: 51 items
Cluster 1: 4 items
Cluster 2: 13 items
Total number of items: 68

```

Gambar 4. Jumlah Cluster

Gambar diatas menunjukkan hasil akhir dari proses K-Means Clustering, dimana data telah dikelompokkan ke dalam tiga cluster berdasarkan data penjualan. C0 memiliki 51 produk yang dikategorikan sebagai Kurang Laris, menunjukkan bahawa sebagian besar produk memiliki tingkat penjualan rendah. C1 hanya terdiri dari 4 produk yang masuk ke dalam kategori Sangat Laris, yang berarti produk ini memiliki tingkat penjualan Tertinggi. Sedangkan C2 terdiri dari 13 produk yang termasuk dalam kategori Laris, yang menunjukkan produk dengan tingkat penjualan sedang. Secara keseluruhan total jumlah produk yang telah dikelompokkan dalam proses ini adalah 68, sesuai dengan total data yang dianalisis. Kemudian untuk melihat berapa nilai akurasi yang dihasilkan dapat dilihat pada Gambar 5.



Gambar 5. Nilai Davies-Bouldin Index

Gambar diatas menunjukkan hasil evaluasi klasterisasi menggunakan Davies-Bouldin Index (DBI) pada aplikasi RapidMiner. Hasil perhitungan DBI yang dihasilkan adalah sebesar 0.374, yang menunjukkan tingkat pemisahan dan kesamaan antar klaster. Semakin kecil nilai Davies-Bouldin Index, semakin baik kualitas klasterisasi, karena klaster yang dihasilkan lebih terpisah dengan tingkat keseragaman yang tinggi di dalamnya.

3.2 Analisa Hasil Menggunakan Rapidminer

Analisa hasil menggunakan RapidMiner yang bertujuan untuk membandingkan hasil perhitungan secara manual menggunakan *Microsoft Excel* dengan *Software RapidMiner*, hasil yang diperoleh dari perhitungan manual dengan *software RapidMiner* didapatkan keanggotaan *cluster* yang sama. Kemudian yang berbeda adalah nama *cluster* nya. Perbandingan perbedaan nama *cluster* yang didapatkan dapat dilihat pada Tabel 9.

Tabel 9. Perbandingan Hasil Pengujian

No	Pengujian Manual Ms. Excel	Pengujian Aplikasi RapidMiner	Jumlah Anggota Cluster
1	Clusters 0	Cluster 0	51
2	Cluster 1	Cluster 2	13
3	Cluster 2	Cluster 1	4

Pada tabel 9 dapat dilihat bahwa jumlah anggota dalam masing-masing cluster tetap sama antara kedua metode. Cluster 0 memiliki 51 anggota, Cluster 1 memiliki 13 anggota dan Cluster 2 memiliki 4 anggota. Namun, terdapat perbedaan dalam penamaan cluster antara kedua metode. Dalam pengujian manual menggunakan Microsoft Excel Cluster 1 berisi 13 anggota, sedangkan pada dalam hasil RapidMiner, cluster dengan jumlah anggota yang sama diberi nama Cluster 2. Begitu juga dengan Cluster 2 dalam perhitungan manual, yang dalam RapidMiner menjadi Cluster 1.

4. Kesimpulan

Penelitian ini membuktikan bahwa penerapan algoritma K-Means Clustering dapat secara efektif mengelompokkan produk liquid vape berdasarkan pola penjualannya. Dengan menggunakan data transaksi dari Vaporate Store dalam rentang waktu Juni hingga November 2024, produk berhasil diklasifikasikan ke dalam tiga kategori utama: kurang laris, laris, dan sangat laris. Hasil analisis menunjukkan bahwa 51 produk termasuk dalam kelompok kurang laris, 13 produk tergolong laris, dan 4 produk berada dalam kategori sangat laris. Kemudian nilai DBI yang dihasilkan dari klasterisasi tersebut sebesar 0.374. Dengan adanya klasifikasi produk yang jelas, toko dapat merancang strategi manajemen stok yang lebih efisien serta meningkatkan efektivitas operasional.

Daftar Rujukan

- [1] M. F. Haryanti *et al.*, "Pengaruh Data Mining, Strategi Perusahaan Terhadap Laporan Kinerja Perusahaan," *J. Manaj. dan Bisnis*, vol. 3, no. 1, pp. 71–90, 2024.
- [2] S. Nabilah, "Pengaruh Penggunaan Teknologi Big Data dalam Bisnis Retail Terhadap Keputusan Konsumen," *WriteBox*, pp. 1–7, 2023, [Online]. Available: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=ApXhns0AAAAJ&citation_for_view=ApXhns0AAAAJ:9yKSN-GCB0IC
- [3] Andi Syahrul Ramdana, Kusri, and E. Pramono, "Penerapan Algoritma K-Means Untuk Manajemen

- Persediaan Di Perpustakaan,” *J. Inform. Teknol. dan Sains*, vol. 6, no. 1, pp. 109–114, 2024, doi: 10.51401/jinteks.v6i1.3911.
- [4] M. D. A. Syahputra, H. Santoso, and F. H. Sibarani, “Implementasi Sistem Pengelolaan Persediaan dengan Algoritma FIFO Pada Gudang Sparepart Sepeda Motor,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 167–176, 2024, doi: 10.47233/jteksis.v6i1.1126.
- [5] F. P. Dewi, P. S. Aryni, and Y. Umidah, “Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.
- [6] H. Supriyanto *et al.*, “Klasterisasi Data Obat Farmasi Berdasarkan Jumlah Persediaan Clustering Pharmaceutical Drug Data Based on Total Inventory Using the K-Means Method,” vol. 13, no. November, pp. 361–369, 2024, doi: 10.34148/teknika.v13i3.987.
- [7] M. A. Sundari, R. Pane, and R. Rohani, “Data Mining Clustering Korban Kejahatan Pelecehan Seksual dengan Kekerasan Berdasarkan Provinsi Menggunakan Metode AHC,” *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3499.
- [8] D. Murni, B. Efendi, and N. Rahmadani, “Implementation of Employee Discipline Clustering At Gotting Sidodadi Village Office Bandar Pasir Mandoge Using K-Means Algorithm,” *J. Tek. Inform.*, vol. 3, no. 2, pp. 295–304, 2022, [Online]. Available: <https://doi.org/10.20884/1.jutif.2022.3.2.236>
- [9] D. Astuti and Muqorobin, “Optimasi Metode K - Means Clustering untuk Pengelompokan Obat Di Puskesmas Mertoyudan I Magelang Optimization of K-Means Clustering Method for Drug Grouping at,” vol. 13, pp. 2144–2160, 2024.
- [10] M. Djaka Permana, A. Lia Hananto, E. Novalia, B. Huda, and T. Paryono, “Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means,” *J. KomtekInfo*, vol. 10, pp. 15–20, 2023, doi: 10.35134/komtekinfo.v10i1.332.
- [11] R. Fikria and S. Sriani, “Analisis Metode K-Means Clustering Dalam Pengelompokan Penjualan Sembako,” *J. Inf. Syst. Res.*, vol. 5, no. 4, pp. 1464–1471, 2024, doi: 10.47065/josh.v5i4.5699.
- [12] G. Triyandana, L. A. Putri, and Y. Umidah, “Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means,” *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 40–46, 2022, doi: 10.30871/jaic.v6i1.3824.
- [13] A. W. Aranski, S. Astiti, and R. A. Putra, “Pengaplikasian Data Mining Dalam Mengelompokan Data Penerima Bantuan Subsidi Rumah dengan Menggunakan Metode K-Means Clustering,” vol. 6, no. 1, pp. 480–489, 2024, doi: 10.47065/bits.v6i1.5366.
- [14] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [15] I. Indra, N. Nur, M. Iqram, and N. Inayah, “Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting,” *INOVTEK Polbeng - Seri Inform.*, vol. 8, no. 2, p. 356, 2023, doi: 10.35314/isi.v8i2.3612.