Progresif: Jurnal Ilmiah Komputer
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0877
p-ISSN: 0216-3284

Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode *K-Means Clustering*

Nova Mutiara[1*] Muhammad Husni Rifqo[2]
Teknik Informatika, Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia
*e-mail *Corresponding Author*: novamutiara0411@gmail.com

Abstract

The purpose of this study is to evaluate the extent to which product clustering results influence consumer satisfaction. Excess inventory can lead to overcrowded and inefficient storage, especially since food, beverages, and other products have expiration dates. Currently, Toko Habib still manages inventory manually, which is time-inefficient and prone to errors. To address this issue, data mining techniques are employed. The technique used in this study is K-means clustering, one of the most popular algorithms due to its ease of implementation. This analysis utilizes the K-means clustering algorithm to categorize data in order to identify fast-selling and slow-selling products, while also preventing product overstock in the warehouse. Clustering is one of the functionalities of data mining, where the clustering algorithm groups a set of data into specific clusters. After clustering, three clusters were selected as the initial centroids. The final results showed that 14 items were highly popular, 42 items were popular, and 160 items were less popular. With these results, Toko Habib can implement policies to maintain the loyalty of potential customers and manage products effectively.

Keywords: *Data Mining, K-Means Algorithm; Cluster Analysis*; Toko Habib

Abstrak

Penelitian ini bertujuan untuk mengevaluasi sejauh mana hasil pengelompokan produk mempengaruhi kepuasan kebutuhan konsumen, Persediaan yang berlebih dapat menyebabkan gudang menjadi penuh sesak dan tidak efisien, terutama karena makanan, minuman, dan produk lainnya mempunyai tanggal kadaluwarsa. Saat ini Toko Habib masih melakukan pengelolaan persediaan secara manual sehingga tidak efisien waktu dan rawan kesalahan. Untuk mengatasi permasalahan tersebut digunakan adalah *K-means clustering* yang merupakan salah satu algoritma yang populer karena mudah diimplementasikan. Analisis ini menggunakan algoritma *clustering K-means* yang dapat mengelompokkan data untuk mengetahui produk laris dan tidak laris, dan juga mencegah penumpukan produk di gudang. *Clustering* merupakan salah satu fungsionalitas data mining, algoritma *clustering* merupakan algoritma pengelompokkan sejumlah data menjadi kelompok–kelompok data tertentu (*cluster*). Setelah pengelompokan, dipilih tiga *cluster* sebagai *centroid* awal. Hasil akhirnya menunjukkan 14 barang sangat laris, 42 barang laris, dan 160 barang kurang laris. Dengan hasil ini, Toko Habib dapat menerapkan kebijakan untuk mempertahankan loyalitas pelanggan potensial dan mengelola produk secara efektif.

Kata Kunci*:* *Data Mining; Algoritma K-Means; Clustering; Toko Habib*

1. Pendahuluan

Penelitian ini penting karena berkaitan dengan mendukung efisiensi pasokan dan manajemen produksi di sektor industri. Dengan meningkatkan akurasi perkiraan pasokan, penelitian ini akan membantu mengurangi biaya dan meningkatkan produktivitas, yang merupakan hal penting dalam persaingan global. Selain itu, penelitian ini mengisi kesenjangan dalam literatur yang ada mengenai metode dan parameter prediksi tawaran dengan mengusulkan atau memodifikasi pendekatan yang ada untuk meningkatkan produk penjualan.Dan dapat dilihat dari perkembangan Teknologi yang semakin pesat telah menghasilkan berbagai macam perangkat lunak yang sangat membantu pengguna komputer dalam menyelesaikan pekerjaan mereka. Penerapan sistem informasi di berbagai aspek kehidupan saat ini telah mengubah cara pandang masyarakat terhadap peningkatan mutu

pekerjaan mereka. Saat ini, banyak data yang tidak diolah untuk menghasilkan informasi atau pengetahuan yang dapat mendukung pengambilan keputusan. Banyak aplikasi yang digunakan untuk menciptakan sistem informasi yang cerdas, salah satunya adalah data mining, yang berperan penting dalam menghasilkan informasi berkualitas tinggi[1].

Banyak perusahaan saat ini menghadapi tantangan dalam memprediksi kebutuhan pasokan secara akurat, yang sering kali mengakibatkan sumber daya terbuang sia-sia, peningkatan biaya, dan ketidakmampuan merespons perubahan permintaan pasar secara efektif. Meskipun berbagai metode prediksi telah dikembangkan, namun masih banyak yang mempunyai keterbatasan oleh karena itu Algoritma *K-Means* efektif untuk beberapa jenis dokumen. Penelitian ini menggunakan analisis data mining dengan metode *clustering K-Means*. Teknik ini memungkinkan pengelompokan data ke dalam klaster berdasarkan kesamaan karakteristiknya, di mana data serupa ditempatkan dalam klaster yang sama, sementara data yang berbeda ditempatkan dalam klaster yang berbeda [2]. *K-means* merupakan salah satu Algoritma yang sangat populer digunakan karena efektivitas dan efisiensinya. Hal ini disebabkan *k-means* sangat mudah dipahami dan dari segi waktu, proses komputasinya cukup cepat [3]. *K-means clustering* merupakan salah satu teknik analisis data atau data mining yang melakukan pemodelan tanpa supervisi tertentu yang suatu metode dan pengelompokan data dengan menggunakan sistem partisi[4]. *Clustering* merupakan salah satu bagian dari Teknik data mining yaitu sekumpulan objek yang mempunyai kesamaan diantara anggotanya dan memiliki ketidaksamaan dengan objek lain pada *cluster* lainnya, dengan kata lain *cluster* adalah sekumpulan objek yang digabung Bersama karna kesamaannya atau kedekatannya[5].

Salah satu Data merupakan kumpulan fakta atau informasi yang diorganisasikan ke dalam suatu bentuk yang dapat diolah atau dimanipulasi. Data digunakan untuk berbagai tujuan seperti analisis penjualan, manajemen inventaris, pelacakan pelanggan, keuangan, dan pengambilan keputusan strategis, dan memainkan peran penting dalam bisnis. Saat ini belum ada metode standar untuk toko habib ketersediaan produk hanya berdasarkan konfirmasi persediaan produk yang ada setelah produk habis, kami akan mengisi kembali stok di toko,cara yang digunakan pada toko ini kurang efektif karena suatu saat toko tersebut membutuhkan produk dalam jumlah besar dan tidak memiliki stok yang cukup akan mengecewakan pelanggan, oleh karena itu diperlukan suatu sistem untuk memantau persediaan produk yang paling laris atau paling sedikit terjual sehingga toko dapat menerapkan teknik data mining dengan teknologi clustering *K-Means* untuk memaksimalkan kepuasan pelanggan [6].

Tujuan dari penelitian ini adalah untuk mengembangkan dan menguji model peramalan pasokan yang lebih akurat dan adaptif yang dapat mengatasi tantangan peramalan permintaan pasokan dalam lingkungan yang dinamis. Penelitian ini juga bertujuan untuk mengidentifikasi dan menganalisis parameter utama yang mempengaruhi keakuratan perkiraan dan memberikan rekomendasi praktis bagi perusahaan untuk menerapkan model perkiraan yang lebih efektif dan efisien.Data penjualan yang ada di toko Habib diolah atau dianalisis untuk mengetahui kisaran trend konsumen di setiap tujuan pemasaran produk ditinjau dari faktor minat. Dengan mengolah data tersebut maka dapat ditentukan pola konsumsi produk Habib Shop secara umum[7]. Data mining melibatkan pengumpulan dan pemanfaatan data historis untuk mengidentifikasi pola-pola tertentu. Pola atau hubungan dalam sejumlah data berukuran besar. Hal ini mencakup analisis statistik dan teknik lain untuk mengidentifikasi pola yang dapat digunakan untuk meningkatkan proses pengambilan keputusan. Data yang tersedia dapat digunakan sebagai sistem pengambilan keputusan solusi penjualan dan penunjang infrastruktur di bidang teknologi, hal inilah yang menjadi alasan munculnya teknologi data mining [8]. Tujuan dari penelitian ini adalah untuk membantu toko habib menentukan produk jual yang termasuk dalam kategori sangat laris, laris dan kurang laris.

2. Tinjauan Pustaka

Penelitian terdahulu yang dilakukan oleh Gunawan dkk. (2014) merupakan salah satu studi penting dalam penerapan data mining, khususnya metode *K-Means clustering*, untuk mendukung pengambilan keputusan dalam bidang pemasaran produk. Studi ini menekankan bagaimana teknologi data mining dapat digunakan secara efektif untuk menggali wawasan dari data penjualan yang besar dan beragam, yang sering kali sulit dianalisis secara manual. Dalam dunia bisnis yang semakin kompetitif, perusahaan perlu mengadopsi pendekatan yang lebih canggih untuk memahami perilaku pasar dan membuat keputusan yang tepat waktu dan akurat. Di sinilah peran data mining menjadi sangat penting. Data mining memungkinkan organisasi

untuk mengidentifikasi pola tersembunyi dalam data yang besar dan tidak terstruktur, serta mengubahnya menjadi informasi yang berguna untuk pengambilan keputusan strategis[9].

Penelitian [10] berfokus pada penerapan metode algoritma *K-Means clustering*, yang digunakan untuk mengelompokkan data menjadi beberapa cluster berdasarkan kemiripan karakteristik. Penelitian ini menggunakan dua alat untuk menerapkan algoritma tersebut: *Microsoft Excel* 2010 dan sebuah aplikasi khusus untuk *K-Means clustering* yang dirancang dan dikembangkan oleh peneliti sendiri. Peneliti selanjutnya disarankan untuk meningkatkan efisiensi aplikasi *K-Means* yang digunakan. Ini bisa dilakukan dengan mengoptimalkan algoritma, memanfaatkan teknik komputasi paralel, atau menggunakan platform yang lebih kuat secara komputasional. Alternatif lainnya, mereka bisa mempertimbangkan penggunaan bahasa pemrograman yang lebih efisien dibandingkan *Excel*.

Pada penelitian ini memfokuskan pengelompokan barang dagangan di sebuah perusahaan ritel menggunakan K-Means untuk mengoptimalkan manajemen stok. Hasilnya menunjukkan pengurangan dalam *overstock* dan *stockout,* yang berkontribusi pada penghematan biaya penyimpanan. Penggunaan algoritma *K-Means* untuk mengelompokkan persediaan barang merupakan pendekatan yang umum dilakukan untuk meningkatkan efisiensi manajemen inventaris. K-Means digunakan untuk mengidentifikasi kelompok-kelompok produk berdasarkan karakteristik tertentu seperti tingkat permintaan, volume penjualan, dan frekuensi restocking. Keberhasilan bisnis jangka panjangtidak hanya bergantung pada penjualan jangka pendek tetapi juga pada kemampuan untuk secara konsisten memenuhi kebutuhan dan persyaratan pelanggan. Hal ini memastikan bahwa pendapatan dari penjualan produk atau penyediaan layanan stabil dan berkelanjutan, yang pada akhirnya menguntungkan penjual dan pembeli[11].

3. Metodologi

3.1 Data dan Parameter

Dalam penelitian ini peneliti menggunakan data mining dengan menggunakan algoritma *K-means*, data minat pelanggan terhadap toko habib dibagi menjadi tiga kelompok untuk mengelola persediaan barang, yaitu: (C1) Produk yang terjual sangat laris, (C2) produk yang terjual laris, (C3) produk yang tidak laris.Tujuannya adalah agar objek-objek yang dikelompokkan menjadi ukuran kedekatan atau hubungannya satu sama lain, dibandingkan dengan kelompok lain[12]. Data mining digunakan juga untuk mengolah pemrosesan data ke dalam jumlah besar yang dibutuhkan database untuk menghasilkan informasi baru yang berguna untuk strategi bisnis[13]. Data yang digunakan pada toko habib yaitu data pada bulan January – February terdapat 216 produk.

Tabel 1. Data barang Toko Habib

NO awal	Nama Barang	satuan	stok awal	Jumlah Terjual	stok Akhir
1	Beras Premium	Kg	150	127	23
2	Beras Premium	Kg	100	75	25
3.	Beras Murah	Kg	100	90	10
4.	Minyak Rose Brand	Box	120	100	20
5.	Minyak Kita	Box	155	70	85
6	Minyak Goreng Kelapa	Box	40	7	33
7.	Minyak Grandco	Box	55	39	16
8.	Minyak sovia	Box	47	23	24
9.	Minyak Bimoli	Box	70	40	30
10.	Gula Pasir	Kg	75	32	43
11	Gula Merah	Kg	30	15	15
12	Susu UHT	Box	45	10	35
13	Susu Kental Manis	Box	30	12	18
14	Susu Bubuk	pack	35	12	23
15	Frisian Flag Kaleng	Box	53	26	27
...	...	...	....	...	...
216	Hydro coco	pack	27	5	22

3.2 Algoritma *K-Means*

K-Means merupakan salah satu solusi untuk pengelompokan berulang dimana algoritma *K-Means* secara acak menentukan nilai *cluster* (K). Nilai ini untuk sementara menjadi pusat pengelompokan dan biasa disebut center centroid atau mean.Kemudian menghitung jarak seluruh data yang ada ke setiap centroid menggunakan rumus *Euclidean* hingga ditemukan jarak terdekat dari semua data berdasarkan kedekatannya dengan centroid [14]. K-means berusaha untuk mengelompokkan data yang berbeda ditempatkan dalam kelompok lainnya. Metode ini memungkinkan analisis yang lebih terfokus terhadap setiap kelompok data berdasarkan kesamaan atau perbedaan karakteristik yang dimiliki berbeda ditempatkan dalam kelompok lain [15].

Algoritma *K-Means* memiliki beberapa aturan yaitu:

1. Tentukan jumlah *cluste*r K yang digunakan untuk mempartisi dataset.
2. Pilih secara acak titik awal K sebagai pusat cluster
3. Temukan pusat *cluster* terdekat untuk semua data Anda.Oleh karena itu, setiap pusat *cluster* mempunyai subset dari dataset dan mewakili sebagian dari dataset tersebut, sehingga membentuk cluster K: C1,2, C3,…., CK .
4. Untuk setiap cluster K, cari pusat cluster baru dan perbarui posisi pusat *cluster* ke nilai baru.
5. Ulangi langkah 3 dan 4 hingga posisi pusat cluster berhenti berubah atau selesai algoritma *K-means* dirumuskan.

Rumus algoritma k-means:

$$D(p,c)_n = \sqrt[2]{\sum_{i=0}^{n}(p_i - c_i)^2|} \quad (1)$$

Pada gambar di atas terlihat rumus algoritma *K-means* yaitu (P) adalah datanya, (C) adalah pusat massanya, (N) adalah kumpulan datanya, dan (i) adalah iterasinya. Algoritma K-means pada dasarnya melakukan dua proses: menemukan posisi centroid untuk setiap kelompok, dan menentukan anggota untuk setiap kelompok tersebut.

Cara kerja algoritma *K-Means*:

1. Tentukan K sebagai banyaknya cluster yang ingin dibentuk
2. Hasilkan secara acak K *centroid* (pusat *cluster*) pertama
3. Hitung jarak semua data ke masing-masing centroid
4. *Centroid* terdekat dari seluruh data adalah.
5. Menentukan letak centroid baru dengan menghitung rata-rata data pada centroid yang sama.

3.3 Metode analisis data

Metode analisis data merupakan pendekatan sistematis yang digunakan untuk memodelkan data dengan tujuan menemukan informasi yang berguna. Pada tahap ini peneliti menemukan dan menganalisis hasil dari perhitungan *K-means clustering* baik itu hasil dari perhitungan secara manual maupun dengan menggunakan alat bantu (tools data mining yaitu dengan *Rapid Miner*). Kemudian hasil yang didapat bisa dijadikan masukan bagi pemilik toko habib karena bisa memudahkan dalam pengelolaan stok barang.

4. Hasil dan Pembahasan

4.1 Implementasi *K-Means*

Data yang digunakan dalam penelitian ini meliputi data penjualan toko Habib bulan Januari dan Februari 2024.Penelitian ini mencakup 216 produk dengan enam atribut yaitu: kuantitas, nama barang, satuan ukuran, persediaan awal, jumlah terjual, dan persediaan akhir, Data ini disajikan dalam format Tabel 1 dari soft file dokumentasi yang diterima dari Toko Habib.

Tabel 2. Data stok barang Toko Habib

NO awal	Nama Barang	satuan	stok awal	Jumlah Terjual	stok Akhir
1	Beras Premium	Kg	150	127	23
2	Beras Premium	Kg	100	75	25
3.	Beras Murah	Kg	100	90	10
4.	Minyak Rose Brand	Box	120	100	20
5.	Minyak Kita	Box	155	70	85
6	Minyak Goreng Kelapa	Box	40	7	33
7.	Minyak Grandco	Box	55	39	16
8.	Minyak sovia	Box	47	23	24
9.	Minyak Bimoli	Box	70	40	30
10.	Gula Pasir	Kg	75	32	43
11	Gula Merah	Kg	30	15	15
12	Susu UHT	Box	45	10	35
13	Susu Kental Manis	Box	30	12	18
14	Susu Bubuk	pack	35	12	23
15	Frisian Flag Kaleng	Box	53	26	27
...	...	...	....	...	...
216	Hydro coco	pack	27	5	22

Berdasarkan Tabel 1, data ini berasal dari operasional harian toko Habib dan mencakup informasi penjualan produk bulan Januari dan Februari. Data primer ini merupakan data mentah yang belum diolah. Untuk menerapkan teknik *clustering K-Means*, anda perlu mengurutkan data berdasarkan variabel tertentu. Penelitian ini menggunakan data terbaru dari toko Habib yaitu data penjualan bulan Januari dan Februari 2024 yang telah diolah menjadi format yang lebih sederhana. Berikut data yang sudah disederhanakan:

Tabel 3. Tabel Data Variabel

NO	Nama Barang	satuan	stok awal	stok akhir
1	Beras Premium	Kg	150	23
2	Beras Premium	Kg	100	25
3.	Beras Murah	Kg	100	10
4.	Minyak Rose Brand	Box	120	20
5.	Minyak Kita	Box	155	85
6	Minyak Goreng Kelapa	Box	40	33
7.	Minyak Grandco	Box	55	16
8.	Minyak sovia	Box	47	24
9.	Minyak Bimoli	Box	70	30
10.	Gula Pasir	Kg	75	43
11	Gula Merah	Kg	30	15
12	Susu UHT	Box	4	35
13	Susu Kental Manis	Box	30	18
14	Susu Bubuk	pack	35	23
15	Frisian Flag Kaleng	Box	53	27
...	...	...	....	...
216	Hydro coco	pack	27	22

Berdasarkan tabel 3 yaitu data primer yang telah diolah menjadi data yang lebih sederhana sehingga dapat diolah berdasarkan jenis variabelnya menggunakan *K-means clstering* dengan tiga *cluster* Penerapan algoritma *K-means clustering* dalam pengelompokan penjualan produk dapat dijelaskan sebagai berikut;

Menentukan nilai centroid awal pada Iterasi 1 Penentuan nilai *centroid* awal pada Iterasi 1 ditentukan secara acak dari data yang ada. Data yang dikumpulkan kali ini adalah data ke-5, data ke-123, dan data ke-6.

Tabel.3 Penentuan Nilai *Centroid* Iterasi 1

No	Cluster	Stok Awal	Stok Akhir
1	C1 data Ke 5	155	85
2	C2 data Ke 123	90	31
3	C3 data Ke 6	40	33

Pada Tabel 3, peneliti memilih tiga data berdasarkan tiga cluster dari seluruh variabel yang digunakan dan data yang terpilih kemudian dihitung menggunakan rumus algoritma *K-means*, dimulai dari data terbesar hingga data terkecil untuk mempermudah perhitungan.
Berikut perhitungan jarak data 1 ke pusat cluster adalah;

D (1,1) = $\sqrt{(150-155)^2+(23-85)^2}$ = 62,2013
D (1,2) = $\sqrt{(150-90)^2+(23-31)^2}$ = 60,531
D (1,3) = $\sqrt{(150-40)^2+(23-33)^2}$ = 110,454

Demikian pula, hitung jarak dari data kedua ke *cluster* data pusat, bandingkan antara ketiga *cluster*, dan pilih nilai minimum. Jika Anda menemukan nilai minimum, Anda dapat mengelompokkan nilai-nilai tersebut ke dalam cluster ini. Berikut hasil perhitungan *cluster* pada iterasi 1:

Tabel.4 Hasil Perhitungan Iterasi ke 1

NO	Nama Barang	satuan	stok awal	stok akhir	C1	C2	C3	Jarak	Cluster
1	Beras Premium	Kg	150	23	62,3013	60,531	110,454	60,5309838	2
2	Beras Premium	Kg	100	25	81,3941	11,6619	60,531	11,66190739	2
3.	Beras Murah	Kg	100	10	93,0054	23,2594	64,2573	23,2594067	2
4.	Minyak Rose Brand	Box	120	20	73,8241	31,9531	81,0494	31,95309062	2
5.	Minyak Kita	Box	155	85	0	84,5044	126,21	0	1
6	Minyak Goreng Kelapa	Box	40	33	126,21	50,04	0	0	3
7.	Minyak Grandco	Box	55	16	121,495	38,0789	22,6716	22,6715681	3
8.	Minyak sovia	Box	47	24	124,036	43,566	11,4018	11,40175425	3
9.	Minyak Bimoli	Box	70	30	101,242	20,025	30,1496	20,02498439	2
10.	Gula Pasir	Kg	75	43	90,3549	19,2094	36,4005	19,20937271	2
11	Gula Merah	Kg	30	15	143,265	62,0967	20,5913	20,59126028	2
12	Susu UHT	Box	45	35	120,83	45,174	5,38516	5,385164807	3
...	...	...	....	...	....	....	....	....	...
216	Hydro coco	pack	27	22	142,664	63,6396	17,0294	17,02938637	3

Berdasarkan Tabel 4 di atas, ditentukan jarak terpendek dan data jarak terpendek dikelompokkan ke dalam *cluster* yaitu, *Cluster* 1 berisi 11 data, cluster 2 berisi 44 data, dan *cluster* 3 berisi 161 data. Namun hal ini tidak mungkin dilakukan hanya pada satu ujung data, sehingga langkah selanjutnya dari algoritma *K-means* harus dilakukan. Data hasil pengelompokan iterasi 1 dihitung menggunakan rumus berikut untuk menentukan nilai centroid pada iterasi 2.

$$Ck = \frac{Jumlah\ dari\ nilai\ yang\ masuk\ ke\ dalam\ cluster}{jumlah\ data\ yang\ masuk} \quad \ldots\ldots\ldots (2)$$

Hasilnya, nilai centroid baru pada iterasi kedua adalah sebagai berikut:

Tabel 5. Penentuan Nilai *Centroid* Iterasi 2

No	Cluster	Stok Awal	Stok Akhir
1	C1	172,27	43
2	C2	94,4	29,77
3	C3	33,71	0

Kemudian Tabel 5 menampilkan nilai centroid baru untuk perhitungan iterasi 2 untuk nilai *centroid* Kemudian, langkah-langkah sebelumnya diulang: setelah nilai *centroid* baru ditemukan, penghitungan jarak dari langkah sebelumnya dilakukan kembali untuk menempatkan data ke dalam *cluster* yang sesuai. Proses ini diulang sampai data di setiap *cluster* sama persis dengan data sebelumnya. Jika langkah ini diulang dengan metode yang sama, sampai data dalam suatu klaster sama persis seperti data sebelumnya, Jika data tidak berubah posisinya dalam klaster, maka penghitungan nilai *centroid* dapat dihentikan. Pada perhitungan iterasi kedua posisi *cluster* pada iterasi 2 mengalami perubahan cluster. Berikut ini adalah hasil perhitungan iterasi 2:

Tabel.6 Hasil Perhitungan Iterasi 2

No	Nama Barang	satuan	stok Awal	stok akhir	C1	C2	C3	Pndktan	Cluster
1.	Beras Premium	kg	150	23	29,93247	56,01065	118,5427	29,93247233	1
2.	Beras Medium	kg	100	25	74,47787	7,356147	70,84747	7,356147089	2
3.	Beras Murah	kg	100	10	79,4478	20,54782	67,04002	20,54781984	2
4.	Minyak Rose Brand	box	120	20	57,1065	27,40097	88,57745	27,40096531	2
5.	Minyak Kita	box	155	85	45,41203	81,99215	148,109	45,41203475	1
6.	Minyak Goreng Kelapa	box	40	33	132,6475	54,49581	33,59411	33,59410811	3
7.	Minyak Grandco	box	55	16	120,3381	41,73695	26,63201	26,63201269	3
8.	Minyak Sovia	box	47	24	126,7027	47,7499	27,473	27,43399533	3
9.	Minyak Bimoli	box	70	30	103,0929	24,40108	47,08465	24,40108399	2
10.	Gula Pasir	kg	75	43	97,27	23,48176	59,61429	23,48175675	2
11.	Gula Merah	kg	30	15	144,9991	66,07203	15,45199	15,4519934	3
12.	Susu UHT	box	45	35	127,5212	49,67608	36,77589	36,77586301	3
...	....	....	....	....	....	.....	.....	.....	...
216.	Hydro Coco	pack	27	22	146,78	67,8464	23,0005	23,0005239	3

Pada tabel perhitungan iterasi ke 2 terdapat perubahan pada *cluster* yaitu antara iterasi 1 dan iterasi 2, jadi kita harus melakukan perhitungan kembali seperti langkah pertama yaitu penentuan nilai *centroid* baru untuk iterasi ke 3. Berikut adalah nilai *centroid* baru untuk perhitungan iterasi ke 3:

Tabel.7 Penentuan Nilai *Centroid* Iterasi 3

No	Cluster	Stok Awal	Stok Akhir
1	C1	167,5	36,64
2	C2	89,97	31,87
3	C3	33,18	15,62

Pada tabel 7 kita sudah menetukan centroid baru untuk perhitungan pada iterasi 3, kemudian untuk melakukan perhitungan iterasi ke 3 kita menggunakan rumus sesuai dengan langkah sebelumnya. Jika penempatan cluster tidak berubah antara iterasi kedua dan iterasi ketiga, maka kita dapat menghentikan penghitungan pada iterasi tersebut. Pada penelitian ini2tidak terdeteksi adanya perubahan *cluster* pada iterasi kedua dan ketiga, sehingga dapat disimpulkan bahwa iterasi ketiga merupakan hasil akhir dari perhitungan *K-means* secara manual Di bawah ini adalah hasil akhir penghitungan untuk iterasi 3:

Tabel.8 Hasil Perhitungan Iterasi 3

No	Nama Barang	satuan	stok Awal	stok akhir	C1	C2	C3	Pndktan	Cluster
1.	Beras Premium	kg	150	23	22,1878	60,6818	117,053	22,1878255	1
2.	Beras Medium	kg	100	25	68,4963	12,1572	67,4752	12,1572119	2
3.	Beras Murah	kg	100	10	72,5668	24,0603	67,0559	24,0602951	2
4.	Minyak Rose Brand	box	120	20	50,3303	32,2908	86,9304	32,2908315	2
5.	Minyak Kita	box	155	85	49,9494	83,9744	140,192	49,9493704	1
6.	Minyak Goreng Kelapa	box	40	33	127,552	49,9828	18,6702	18,6702116	3
7.	Minyak Grandco	box	55	16	114,378	38,4026	21,8233	21,8233086	3
8.	Minyak Sovia	box	47	24	121,161	43,6848	16,1622	16,1622028	3
9.	Minyak Bimoli	box	70	30	97,7258	20,0574	39,5284	20,0573627	2
10.	Gula Pasir	kg	75	43	92,7184	18,6542	49,9858	18,6541631	2
11.	Gula Merah	kg	30	15	139,192	62,2977	3,23988	3,23987654	3
12.	Susu UHT	box	45	35	122,511	45,0788	22,7001	22,7001498	3
…	….	….	….	….	….	…..	…..	…..	…
216.	Hydro Coco	pack	27	22	141,261	63,7388	8,88239	8,88238707	3

Hasil analisis menggunakan algoritma *K-means* yang dilakukan untuk mengetahui minat pelanggan terhadap toko habib ini diperoleh dari data yang telah dihitung menggunakan *Microsoft Excel*. Perhitungan dilakukan dengan terlebih dahulu membagi data ke dalam beberapa klaster, yaitu transaksi dan jumlah penjualan. Proses iterasi dihentikan ketika hasil rasio memiliki nilai yang sama dengan rasio sebelumnya. Dalam penelitian ini, proses iterasi dilakukan sebanyak tiga kali. Dan pada perhitungan iterasi ke 3 tidak terjadi perubahan lagi pada setiap cluster, maka hasil akhir nya yaitu terdapat 14 barang dengan kategori sangat laris, 40 barang dengan kategori laris, dan 160 barang dengan kategori tidak laris.

4.2 Pembahasan

Penelitian ini menggunakan algoritma *K-Means* untuk mengklasifikasikan produk di toko Habib berdasarkan popularitas penjualan. Hasil penelitian ini menunjukkan bahwa terdapat pengelompokkan dengan tiga kategori: Sangat laris, Laris, dan tidak laris, yang dapat digunakan untuk mengoptimalkan strategi penjualan dan manajemen inventaris. Dalam hal ini, penelitian ini relevan dan menguatkan hasil penelitian sebelumnya yang mengandalkan *K-means* dalam tugas serupa. Secara keseluruhan penelitian ini tidak hanya mengkonfirmasi hasil penelitian sebelumnya tentang efektivitas *K-means* dalam klastering data penjualan tetapi juga menambah konteks baru, yaitu pengelolaan stok barang dan strategi penjualan di toko habib. Hal ini menegaskan bahwa *K-means* adalah alat yang efektif untuk berbagai aplikasi dalam pengelolaan data, baik di sektor Kesehatan maupun retail, memperluas relevansi metode k-means dalam berbagai konteks praktis.

Pada penelitian ini menunjukkan bahwa algoritma *K-means* sangat efektif digunakan dalam mengelompokkan data penjualan untuk mengidentifikasi pola pembelian dan manajemen stok di toko. Peneletian ini juga memperkuat temuan tersebut dengan menambah bukti tentang bagaimana *K-means* dapat diterapkan dalam konteks retail, khususnya untuk memprediksi kategori barang berdasakan perfoma penjualan[16]. Selain itu pengguanaan *K-means* juga dilakukan dalam menganalisis data penjualan produk di toko Sepatu, penelitian ini menunjukkan bahwa metode *K-means* dapat membantu dalam pengambilan Keputusan strategis berbasis data. *K-means* juga sangat membantu pemilik toko memahami pola penjualan dan membuat Keputusan yang lebih tepat [17]. Penelitian lain juga menggunakan metode *K-means* untuk mengelompokkan data Kesehatan, ini menunjukkan bahwa metode *K-means* dapat digunakan secara luas dalam berbagai domain, termasuk Kesehatan, ini memperkuat hasil tersebut dengan menunjukkan fleksibilitas metode *K-means* dalam mengelola data dari sektor yang berbeda dan bagaimana metode ini dapat diterapkan dalam konteks penjualan retail [18].

5. Simpulan

Pada penelitian ini tidak dilakukan uji akurasi untuk menilai seberapa baik metode yang digunakan yaitu *K-means clustering* dalam mengelompokkan data. Hal ini merupakan salah satu keterbatasan penelitian yang dapat mempengaruhi validitas hasil yang diperoleh. Tanpapengujian akurasi, sulit untuk menentukan apakah *cluster* yang terbentuk benar-benar cocok dengan pola pada data yang ada. Berdasarkan penelitian di atas, dapat diambil kesimpulan bahwa data untuk penelitian ini diperoleh dari data transaksi penjualan toko Habib dan data yang digunakan dalam penelitian ini adalah data penjualan pada bulan Januari sampai dengan bulan Februari. Selanjutnya, modelkan dataset yang ada menggunakan *cluster K-means* dan uji hasil pengelompokannya menggunakan kinerja jarak *cluster*. Studi ini mengidentifikasi tiga kelompok yaitu sangat laris, laris dan tidak laris. Hasil data menunjukkan bahwa hasil rekomendasi penelitian ini diperoleh dengan menerapkan algoritma *K-Means* pada data penjualan, 14 item terjual dengan sangat baik, 42 item masuk dalam kategori bestseller, bestseller menjelaskan bahwa terdapat 160 item dalam kategori tersebut. Kategori yang tidak terjual dengan baik. Hal ini memungkinkan pemilik toko Habib untuk menerapkan strategi penjualan dan pembelian kembali berdasarkan produk terlaris mereka.

Pada penelitian ini masih terdapat banyak kekurangan sehingga saran bagi peneliti selanjutnya adalah untuk melakukan pengujian keakuratan metode *clustering* yang digunakan. Pengujian ini dapat dilakukan dengan menggunakan berbagai metrik evaluasi seperti skor siluet dan indeks Davis-Bouldin untuk mengevaluasi kualitas *cluster* yang terbentuk. Dengan cara ini, hasil penelitian akan lebih valid dan dapat berkontribusi lebih besar lagi terhadap pengambilan keputusan strategis berdasarkan data yang dianalisis.

Daftar Referensi

[1] I. Syafrinal and E. L. Febrianti, “Penerapan Algoritma K-Means Pada Aplikasi Data Mining Untuk Menentukan Pola Penjualan (Studi Kasus: Zahra Mart),” *J. Digit*, vol. 13, no. 1, pp. 31-40, 2023, doi: 10.51920/jd.v13i1.320.

[2] M. H. Siregar, “Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan),” *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.

[3] P. Alam Jusia, F. Muhammad Irfan, and S. Dinamika Bangsa Jambi Jl Jend Sudirman Thehok Jambi, “Clustering Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means,” *J. IKRA-ITH Inform.*, vol. 3, no. 3, pp. 75-84, 2019.

[4] A. H. A. N. Karsa and A. R. Hidayat, “Metode Algoritma K-Means Untuk Clustering Data Produk Paling Laku Pada Toko Tono Grosir Plumbon Cirebon,” *Syntax Lit.; J. Ilm. Indones.*, vol. 7, no. 9, pp. 15984–15996, 2024, doi: 10.36418/syntax-literate.v7i9.15144.

[5] I. Safira, R. Salkiawati, and W. Priatna, “Penerapan Algoritma K-Means untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi,” *J. Inform. Inf. Secur.*, vol. 3, no. 1, pp. 99–110, 2022, doi: 10.31599/jiforty.v3i1.1253.

[6] Kasini and N. Hidayati, “Penerapan Data Mining Untuk Clustering Pada Toko Laura Grosir Dan Eceran Menggunakan Algoritma K-Means,” *JUSTER J. Sains dan Terap.*, vol. 2, no. 3, pp. 51–60, 2023, doi: 10.57218/juster.v2i3.990.

[7] Sutrisno, Afriyudi, and Widiyanto, “Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt . Indomarco,” *Penerapan Data Min. Pada Penjualan Menggunakan Metod. Clust.*, vol. Vol.x No.x, no. Data Mining, pp. 1–11, 2013, [Online]. Available: http://eprints.binadarma.ac.id/78/1/Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt. Indomarco Palembang.pdf

[8] M. H. Fakhriza, and K. Umam, “Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Dalam Pengelompokkan,” *JIKA*, Vol. 5, no.1, pp. 8–15, 2021.

[9] A. S. Gunawan, E. M. Sipayung, and Alvin, “Menggunakan Data Mining Dengan K-Means Clustering,” *Semin. Nas. Sist. Inf. Indones.*, September, 2014, pp. 1–6.

[10] G. Gustientiedina, M.H. Adiya, & Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. J*urnal Nasional Teknologi Dan Sistem Informasi*, vol. 5, no. 1, pp. 17-24, 2019.

[11] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, “Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering,” *JATI*

(Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 615–620, 2024, doi: 10.36040/jati.v8i1.8360.
[12] Y. Yulianti, D. Y. Utami, N. Hikmah, and F. N. Hasan, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Mengetahui Minat Customer Di Toko Hijab," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 241–246, 2019, doi: 10.33480/pilar.v15i2.650.
[13] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, pp. 109-118, 2019, doi: 10.30595/juita.v7i2.5529.
[14] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, pp. 522-531, 2022, doi: 10.30865/json.v3i4.4242.
[15] A. Fikri Sallaby, R. Tri Alinse, V. Novita Sari, and T. Ramadani, "Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil PenjualanDi Toko Widya Bengkulu Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil Penjualan Di Toko Widya Bengkulu," *J. Media Infotama*, vol. 18, no. 1, pp. 29-40, 2022.
[16] O. Diana Hidayati and M. Adrian Juniarta Hidayat, "Implementasi Data Mining Menggunakan Algoritma K-Means Untuk Pengelompokkan," *Jurnal Pengembangan Rekayasa Informatika dan Komputer*., *November*, vol. 1, no. 2, pp 3052-9142, 2023.
[17] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)." *JUKANTI (jurnal pendidikan Tek. Inform*.,vol. 5, no. 2, pp 2621-1467, 2022.
[18] A. Pujianti and M. Mulyawan, "Implementasi Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Status Kematian Bayi Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 459–463, 2023, doi: 10.36040/jati.v7i1.6347.