

Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten

Normah¹, Siti Nurajizah², Arinda Salbinda³

¹Teknik Informatika, Universitas Nusa Mandiri

¹e-mail: normah.nor@nusamandiri.ac.id

²Sistem Informasi Akuntansi Kampus Kabupaten Karawang, Universitas Bina Sarana Informatika

²e-mail: siti.snz@bsi.ac.id

³Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri

²e-mail: arindasalbinda206@gmail.com

Diterima	Direvisi	Disetujui
07-05-2021	08-06-2021	22-06-2021

Abstrak - Toko Helai merupakan sebuah toko yang bergerak dalam bidang penjualan fashion hijab, namun demikian dari berbagai jenis pakaian yang dijual tentu tidak semuanya yang laris terjual, ada juga yang kurang laris terjual. Data-data penjualan, pembelian barang maupun pengeluaran tidak terduga pada Toko Helai ini tidak tersusun dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip bagi toko dan tidak dapat dimanfaatkan untuk pengembangan strategi pemasaran. Oleh karena itu perlu diterapkan data mining menggunakan metode K-Means pada Toko Helai. Metode K-Means dapat diterapkan pada Toko Helai untuk menentukan penjualan baju mana yang sangat laris, laris dan kurang laris. Penerapan metode K-Means pada toko Helai, yaitu dengan cara mengelompokkan data stok baju. Kemudian memilih 3 cluster secara acak sebagai centroid awal. Setelah data pada setiap *cluster* tidak berubah-ubah, maka dapat diketahui hasil akhirnya yaitu yang sangat laris ada 11 artikel, yang laris ada 55 artikel dan 34 artikel untuk yang kurang laris. Kemudian Menerapkan metode K-Means pada Rapidminer dilakukan dengan memasukkan data stok produk yaitu stok awal, stok terjual dan stok akhir yang akan menjadi *Database* pada Ms.Excel, data tersebut kemudian dikoneksikan ke dalam *Tools* Rapidminer, dan akan diolah dan dibentuk K-means. Setelah itu, Rapidminer akan menghasilkan produk mana yang sangat laris, laris, dan kurang laris.

Kata Kunci: Data Mining, K-Means, Rapidminer.

Abstract - *The Helai store is a shop that specializes in selling hijab clothing, however, of the various types of clothing sold, of course not all are sold, some are not sold well. Data on sales, purchase of goods or unexpected expenses in the shop the hijab is not arranged properly, so that the data only serves as an archive for the store and can not be utilized for the development of marketing strategies. Therefore, it is necessary to apply data mining using the K-Means method of the Hijab shop. The K-Means method can be applied to the hijab shop to determine which clothing sales are best selling, bestselling and not selling. Application of K-Means method at the shop hijab, that is by grouping the stock data. Then select 3 clusters randomly as the initial centroid. After the data on each cluster is not changing, it can be known the end result is a very bestselling 11 articles, the bestselling 55 articles and 34 articles for the less-selling. Then apply the K-Means method on Rapidminer is done by inserting the stock data of the product that is the initial stock, the stock is sold and the final stock will be the Database in Ms. Excel, the data is then connected to Rapidminer Tools, and will be processed and formed K-means. After that, Rapidminer will produce the best selling, bestselling, and less-selling products.*

Keywords: Data Mining, K-means, Rapidminer.

PENDAHULUAN

1. Latar Belakang

Dalam persaingan dunia bisnis pada saat ini, kita dituntut untuk selalu mengembangkan bisnis agar bertahan dalam persaingan, khususnya dalam

persaingan penjualan menuntut para pengusaha untuk menemukan suatu pola yang dapat meningkatkan penjualan dan pemasaran di perusahaan, salah satunya adalah dengan pemanfaatan data penjualan.”(Siregar, 2018)

Hal ini tidak dimanfaatkan oleh salah satu toko pakaian yaitu Toko Helai. Toko Helai merupakan sebuah toko yang bergerak dalam bidang penjualan Pakaian Muslimah, namun demikian dari berbagai jenis pakaian yang dijual tentu tidak semuanya yang sangat laris terjual, yang laris, serta kurang laris. Data-data penjualan, pembelian barang maupun pengeluaran tidak terduga pada toko Helai ini tidak tersusun dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip bagi toko dan tidak dapat dimanfaatkan untuk pengembangan strategi pemasaran. Seharusnya data yang telah tersedia dapat dijadikan sebagai sistem pengambilan keputusan untuk solusi bisnis serta dukungan infrastruktur di bidang teknologi yang merupakan penyebab munculnya suatu teknologi yaitu data mining.

Data mining yaitu suatu proses penggalian data data dari sebuah informasi yang sangat penting. Data Mining juga merupakan suatu proses untuk menggali pola-pola dari data. Pola-pola itu didapatkan dari berbagai jenis basis data seperti basis data relasional, data warehouse, data transaksi, dan data berorientasi objek. Penggunaan data mining dapat membantu para pebisnis dalam pengambilan keputusan secara cepat dan tepat (Nurajizah, 2019).

Menentukan pola penjualan dengan melihat kecenderungan pembelian konsumen jika dianalisa dan diolah dengan baik maka dapat membantu mengetahui produk mana yang terlaris dan kurang laku terjual, sehingga kontrol stok persediaan dapat dilakukan, dan dapat dijadikan sebagai masukan bagi perusahaan dalam mengembangkan strategi pemasaran (Normah, Rifai, & Sari, 2020).

Output Data Mining dapat digunakan untuk membantu pengambilan keputusan di masa mendatang (Normah, Yulianti, et al., 2020). "K-Means ialah metode Clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih Cluster atau dapat mempunyai tujuan untuk membagi data menjadi beberapa kelompok (Bahar, Pramono, & Sagala, 2016)". Algoritma k-means untuk mengetahui minat customer terhadap produk yang dijual (Yulianti, Utami, Hikmah, & Hasan, 2019).

2. Identifikasi Masalah

1. Bagaimana cara menerapkan metode Data Mining dengan algoritma K-Means untuk menentukan produk yang sangat laris, laris dan yang kurang laris?
2. Bagaimana cara perusahaan tidak mengalami kesulitan untuk mengelompokkan baju apa saja yang dibutuhkan konsumen dan penyimpanan data-data yang kurang efektif?

3. Maksud dan Tujuan

Adapun maksud dari penelitian ini adalah : Mengimplementasikan konsep data mining menggunakan Algoritma K-Means untuk menentukan

produk yang sangat laris, laris dan yang kurang laris sebagai bahan untuk meningkatkan penjualan

Sedangkan tujuan dari penelitian ini adalah membantu Toko Helai dalam menentukan produk penjualan yang termasuk ke dalam kategori sangat laris, laris serta kurang laris.

4. Maksud dan Tujuan

1. Observasi
Penulis melakukan pengamatan secara langsung di bagian Penjualan di toko yang beralamat di Cikuya Kecamatan Solear Tangerang Banten.
2. Wawancara
Penulis memperoleh data-data serta meneliti kebenaran informasi data-data tersebut dengan melakukan tanya jawab secara langsung dengan pemilik toko yaitu Ibu Indah.
3. Studi Pustaka
Melakukan studi kepustakaan untuk pengumpulan data dan informasi yang diperoleh dari perpustakaan, jurnal-jurnal, artikel dan internet yang berkaitan dengan penelitian ini.

METODOLOGI PENELITIAN

Adapun kerangka kerja pada penelitian ini diuraikan sebagai berikut :



Gambar 1. Kerangka Kerja Penelitian
Sumber : Hasil Penelitian (2021)

Tahapan pada penelitian ini dapat dijelaskan sebagai berikut :

1. Identifikasi Masalah
Yaitu melakukan identifikasi masalah yang terdapat dalam toko yaitu untuk mengetahui produk baju mana saja yang sangat laris, laris, dan kurang laris sebagai acuan strategi persediaan stok baju pada toko.
2. Pengumpulan Data
Yaitu mengumpulkan data dengan meminta data stok baju berdasarkan data stok baju periode bulan juli 2019 sampai dengan september 2019 pada toko Helai.
3. Analisa Data
Yaitu menganalisa permasalahan ,yaitu dengan menganalisa data stok baju pada toko Helai.
4. Pengolahan Data
Yaitu dengan mengolah data yang diberikan toko helai, lalu data tersebut diolah dengan metode K-means clustering yang dihitung secara manual untuk mengelompokkan pakaian yang sangat laris, laris dan kurang laris.
5. Pengujian Data
Yaitu dengan melakukan pengujian data yang sudah diolah dengan Rapid Miner.
6. Pengambilan Kesimpulan
Yaitu dengan menarik kesimpulan dari hasil perhitungan dan pengolahan data yang dihitung secara manual dengan data hasil Rapid Miner

HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini adalah data stok baju pada bulan juli 2019 sampai dengan september 2019 pada toko Helai. Data dibawah ini adalah data yang akan diolah, berupa stok awal, stok terjual dan stok akhir:

Tabel 1. Data Stok Toko helai Bulan Juli 2019 – September 2019

No	Artikel	SA	ST	SAK
1	PolyC pink	54	41	13
2	polyC abu	30	25	5
3	PolyC milo	47	40	7
86	Kisana	84	63	21
87	Clay Batwing	34	15	19
90	Cactus Batwing	30	21	9
91	Voal Steelblue	30	17	13
92	Catrice Batwing	33	8	25
93	Tartan Batwing	50	13	37
94	Long Outer	67	29	38
95	Najwa Batwing	44	20	24
96	Rania Batwing	50	37	13
97	Numira Batwing	36	27	9
98	Voal Sahara	20	14	6
99	Abstak Batik	25	1	24
100	Grasia	29	7	22

Sumber : Hasil Penelitian (2021)

Langkah-langkah dalam mengcluster menggunakan metode K-Means adalah sebagai berikut :

1. Tentukan nilai k nya sebagai jumlah kluster yang akan dibentuk.
2. Tentukan Titik pusat awal dari setiap kluster.
3. Hitunglah jarak setiap data input masing – masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) sampai ditemukan jarak yang terdekat dari setiap data dengan centroid.

Berikut adalah persamaan Euclidian Distance :

$$D(x, y) = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2} \dots\dots\dots(1)$$

Keterangan;

D = Jarak

x = Data

y = Centroid

4. Mengklasifikasi data berdasarkan kedekatannya dengan centroid.
5. Hitunglah kembali pusat kluster dengan anggota cluster yang sekarang. Pusat cluster ialah nilai rata-rata dari semua data objek dalam cluster tertentu.
6. Hitung lagi setiap objek memakai pusat kluster yang baru. Jika pusat cluster tidak berubah lagi maka proses klustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat kluster tidak berubah lagi.

Tabel 2. Centroid Awal

Keterangan	Stok Awal	Stok Terjual	Stok Akhir
C1	84	63	21
C2	30	18	12
C3	53	33	20

Sumber : Hasil Pengolahan Data (2021)

Tabel II adalah Menentukan Pusat Awal Cluster secara acak diambil dari data pada Tabel IV.1. Data yang dipilih untuk cluster awal adalah produk ke-86, 56, 77.

Jarak centroid data ke-1 pada cluster 1 (C1) yaitu:

$$D(x1, c1) = \sqrt{(SA_1 - C1_1)^2 + (ST_1 - C1_1)^2 + (SAK_1 - C1_1)^2} = \sqrt{(54 - 84)^2 + (41 - 63)^2 + (13 - 21)^2}$$

$$= 900 + 484 + 6$$

$$= 38,0525$$

Jarak centroid data ke-1 pada cluster 2 (C2) yaitu:

$$D(x1, c2) = \sqrt{(SA_1 - C2_1)^2 + (ST_1 - C2_1)^2 + (SAK_1 - C2_1)^2} = \sqrt{(54 - 30)^2 + (41 - 18)^2 + (13 - 12)^2}$$

$$= 576 + 529 + 1$$

$$= 33,2565$$

Jarak centroid data ke-1 pada cluster 3 (C3) yaitu:

$$D(x1, c3) = \sqrt{(SA_1 - C3_1)^2 + (ST_1 - C3_1)^2 + (SAK_1 - C3_1)^2} = \sqrt{(54 - 53)^2 + (41 - 33)^2 + (13 - 20)^2}$$

$$= 1 + 64 + 49$$

$$= 10,6770$$

Selanjutnya menghitung centroid baru untuk setiap cluster sesuai data yang tergabung pada setiap clusternya.

Tabel 3. Data Cluster 1 Iterasi 1

Artikel	SA	ST	SAK
Pashmina pink	101	50	51
Pashmina coksu	96	63	33
PC black	104	101	3
Zahra Black	112	106	6
Sabrina Pendek	64	61	3
Jasmin White	100	42	58
Magnolia	93	73	20
Bsc Line Marun	86	55	31
Bsc Line Red Velvet	66	64	2
Turkishawl Mocca	90	49	41
Kisana	84	63	21
Jumlah	996	727	269
Rata -Rata	90,5454	66,090	24,4545

Sumber : Hasil Pengolahan Data (2021)

Tabel 3 bahwa cluster 1 pada iterasi pertama menghasilkan 11 data yang tergabung ke dalamnya.

Tabel 4. Data Cluster 2 Iterasi 1

Artikel	SA	ST	SAK
Polyc Abu	30	25	5
Pashmina Green	35	21	14
Pashmina Purple	24	23	1
Pashmina Milo	27	21	6
Pashmina Silver	33	24	9
Pashmina Caramel	35	31	4
Pashmina Vanilla	40	29	11
PC Maroon	36	24	12
PC Yellow	40	8	32
Lily Hijau	25	23	2
Lily Ungu	25	13	12
Aisyah Green	35	25	10

Sumber : Hasil Pengolahan Data (2021)

Tabel 4 bahwa Untuk Cluster 2 terdapat 55 data yang tergabung didalamnya.

Tabel 5. Data Cluster 3 Iterasi 1

Artikel	SA	ST	SAK
Polyc Pink	54	41	13
Polyc Milo	47	40	7
Polyc Mint	50	34	16
Pashmina Peach	48	38	10
Pashminabata	44	31	13
PC Softblue	57	53	4
Lily Pink	45	38	7
Isabela Cerut Pink	44	43	1
Rajut Skirt Biru	58	45	13
Rajut Skirt Black	67	44	23
Sabrina Panjang	50	5	45
PC Softblue	57	53	4
Denim Pants	60	46	14
Mocca Pants	42	31	11
Cornskin Blue	56	36	20

Sumber : Hasil Pengolahan Data (2021)

Tabel 5 bahwa Untuk Cluster 3 terdapat 34 data yang tergabung didalamnya.

Tabel 6. Centroid Baru

Ket	SA	ST	SAK
C1	90,5454545	66,091	24,45
C2	31,0727273	19,436	11,64
C3	51,5	33	20,18

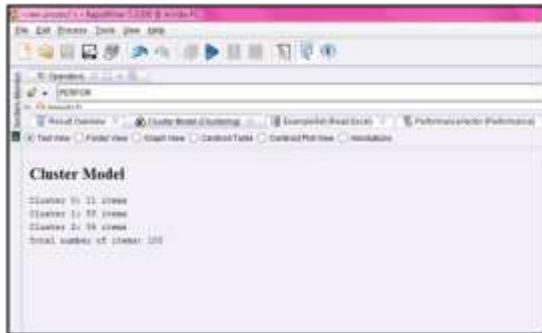
Sumber : Hasil Pengolahan Data (2021)

Tabel 6 adalah centroid baru yang dihasilkan dari rata-rata dari 3 cluster.

Centroid baru ini dihitung menggunakan rumus seperti rumus iterasi per-1 dan karena tidak ada data yang berpindah cluster dan cluster ke- 1 dan ke-2 hasilnya sama, maka proses perhitungan centroid yang baru dihentikan dan berakhir pada iterasi ke-2.

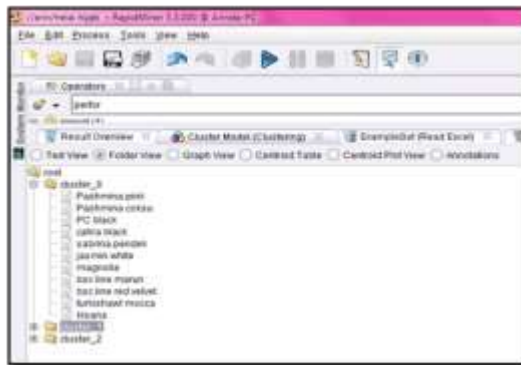
K-MEANS PADA RAPIDMINER

1. Tampilan Hasil Clustering



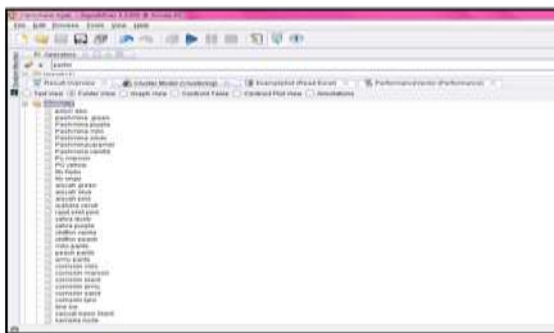
Sumber : Hasil Pengolahan Data (2021)

Gambar 1. Tampilan Hasil Cluster Model
Gambar 1 adalah tampilan hasil cluster model dari data yang diolah.



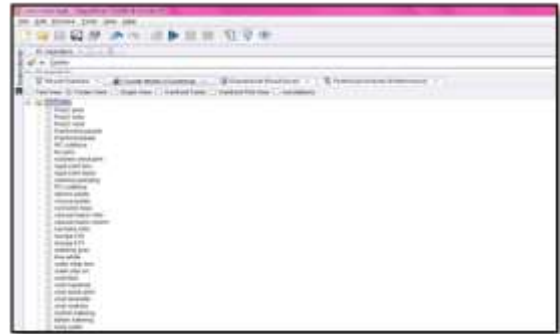
Sumber : Hasil Pengolahan Data (2021)

Gambar 2. Tampilan Hasil Cluster 1
Gambar 2 adalah tampilan hasil cluster 1 yang sudah diolah dari Rapidminer.



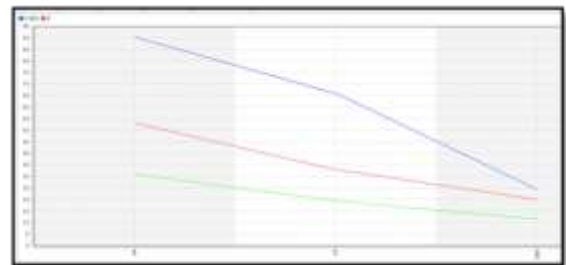
Sumber : Hasil Pengolahan Data (2021)

Gambar 3. Tampilan Hasil Cluster 2
Gambar 3 adalah tampilan hasil dari cluster 2 yang sudah diolah dari Rapidminer.



Sumber : Hasil Pengolahan Data (2021)

Gambar 4. Tampilan Hasil Cluser 3
Gambar 4 adalah tampilan hasil dari cluster 3 yang sudah diolah dari Rapidminer.



Sumber : Hasil Pengolahan Data (2021)

Gambar 5. Hasil Pengolahan Data Pada Rapidminer
Gambar 5 merupakan hasil pengolahan data pada Rapidminer yang ditandai dengan warna biru merupakan hasil dari cluster 1, warna hijau adalah hasil dari cluster 2 dan warna merah merupakan hasil dari cluster 3.

KESIMPULAN

1. Metode K-means dapat diterapkan pada toko Helai untuk menentukan penjualan baju mana yang sangat laris, laris dan kurang laris. Penerapan metode K-Means pada toko Helai, yaitu dengan cara mengelompokkan data stok baju. Kemudian memilih 3 cluster secara acak sebagai centroid awal. Setelah data pada setiap *cluster* tidak berubah-ubah, maka dapat diketahui hasil akhirnya yaitu yang sangat laris ada 11 artikel, yang laris ada 55 artikel dan 34 artikel untuk yang kurang laris.
2. Metode Data Mining menggunakan *K-Means* sangat membantu dan mempermudah toko Helai untuk mengembangkan strategi persediaan stok baju .
3. Menerapkan metode *k-means* pada Rapidminer dilakukan dengan memasukkan data stok produk yaitu stok awal, stok terjual dan stok akhir yang akan menjadi *Database* pada Ms.Excel, data tersebut kemudian dikoneksikan ke dalam *Tools* Rapidminer, dan akan diolah dan dibentuk *K-means*. Setelah itu,

Rapidminer akan menghasilkan produk mana yang sangat laris, laris, dan kurang laris.

REFERENSI

- Bahar, A., Pramono, B., & Sagala, L. H. S. (2016). Penentuan strategi penjualan alat-alat tattoo di studio sonyxtattoo menggunakan metode. *SemanTIK*, 2(2), 75–86.
- Normah, N., Rifai, B., & Sari, P. (2020). Algoritma Apriori Sebagai Solusi Kontrol Persediaan Suku Cadang Mobil PT. Buanasakti Aneka Motor Jakarta. *Paradigma - Jurnal Komputer Dan Informatika*, 22(2), 161–168. <https://doi.org/10.31294/p.v22i2.6530>
- Normah, Yulianti, I., Novianti, D., Winnarto, M. N., Zumarniansyah, A., & Linawati, S. (2020). Comparison of Classification C4.5 Algorithms and Naïve Bayes Classifier in Determining Merchant Acceptance on Sponsorship Program. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012006>
- Nurajizah, S. (2019). Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori. *INOVTEK*, 4(1), 35–44.
- Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, 1(2), 83–91. <https://doi.org/10.36378/jtos.v1i2.24>
- Yulianti, Y., Utami, D. Y., Hikmah, N., & Hasan, F. N. (2019). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Mengetahui Minat Customer Di Toko Hijab. *Jurnal Pilar Nusa Mandiri*, 15(2), 241–246. <https://doi.org/10.33480/pilar.v15i2.650>