

Analisis Metode K-Means Clustering pada Data Penjualan Barang NDT PT. Sandya Lestari

¹Mega Permata Sapani, ²Yossy Veiebrian Fitri Prasmono

^{1,2}Program Studi Sistem Informasi, Universitas Pamulang, Tangerang Selatan

E-mail: ¹dosen03318@unpam.ac.id, ²dosen03319@unpam.ac.id

ABSTRAK

NDT (Non-Destructive Testing) merupakan teknik pengujian material tanpa merusak benda yang diuji, dengan produk utama antara lain Liquid Penetrant Testing dan Magnetic Particle Testing. Penumpukan stok barang yang belum terjual pada PT. Sandya Lestari menjadi pertimbangan penting dalam penyusunan strategi pemasaran dan penjualan. Penelitian ini bertujuan menganalisis data penjualan barang NDT menggunakan metode K-Means Clustering untuk mengelompokkan barang berdasarkan tingkat penjualannya. Data yang digunakan terdiri atas 28 jenis barang dengan atribut Stok Awal (SA), Stok Terjual (ST), dan Stok Akhir (SAK) pada rentang waktu tertentu. Proses clustering dilakukan baik melalui perhitungan manual menggunakan jarak Euclidean maupun dengan bantuan perangkat lunak RapidMiner Studio, terhadap 3 cluster. Hasil penelitian menunjukkan posisi pusat cluster konvergen pada iterasi ke-3, dengan hasil akhir Cluster 0 (C0) berisi 25 item barang kurang laku, Cluster 1 (C1) berisi 2 item barang laku, dan Cluster 2 (C2) berisi 1 item barang paling laku. Nilai validitas cluster yang diukur menggunakan Cluster Distance Performance (Average Within Centroid Distance) sebesar 21.073,928. Hasil perhitungan manual dan pengolahan software menunjukkan kesesuaian, dan mengindikasikan bahwa sebagian besar item (25 dari 28) tergolong kurang laku sehingga berkontribusi terhadap penumpukan stok barang. Temuan ini dapat menjadi dasar penyusunan strategi pemasaran dan pengelolaan persediaan yang lebih efektif bagi perusahaan.

Kata Kunci: *K-Means, Clustering, Data Penjualan, Barang NDT, PT. Sandya Lestari*

ABSTRACT

NDT (Non-Destructive Testing) is a material testing technique that does not damage the object being tested, with its main products including Liquid Penetrant Testing and Magnetic Particle Testing. The accumulation of unsold stock at PT. Sandya Lestari has become an important consideration in formulating marketing and sales strategies. This study aims to analyze NDT sales data using the K-Means Clustering method to group items based on their sales level. The data used consists of 28 items with the attributes of Initial Stock (SA), Sold Stock (ST), and Final Stock (SAK) over a given period. The clustering process was carried out both manually using Euclidean distance and with the help of RapidMiner Studio software, using 3 clusters. The results show that the cluster centers converged at the third iteration, with the final result being Cluster 0 (C0) containing 25 slow-moving items, Cluster 1 (C1) containing 2 moderately-selling items, and Cluster 2 (C2) containing 1 best-selling item. The cluster validity value measured using Cluster Distance Performance (Average Within Centroid Distance) was 21,073.928. The manual calculation and the software processing produced consistent results, indicating that most items (25 of 28) are classified as slow-moving, contributing to stock accumulation. These findings can serve as a basis for developing more effective marketing strategies and inventory management for the company.

Keywords: *K-Means, Clustering, Sales Data, NDT Products, PT. Sandya Lestari*

PENDAHULUAN

NDT (Non-Destructive Testing) merupakan suatu teknik pengujian material tanpa merusak benda yang diuji, yang dilakukan agar material yang sedang digunakan tetap aman dan tidak mengalami kerusakan. Produk NDT di antaranya Liquid Penetrant Testing dan Magnetic Particle Testing. Kemajuan teknologi informasi yang pesat menghasilkan volume data yang besar di berbagai bidang, termasuk sektor ritel. Dalam lima tahun terakhir, rata-rata pertumbuhan jumlah gerai ritel di Indonesia tercatat sekitar 1.167 unit per tahun.

Pengelolaan data penjualan yang baik perlu diubah menjadi informasi yang dapat mendukung pengambilan keputusan bisnis, salah satunya melalui

pendekatan data mining. Data mining bersifat otomatis dan interaktif dalam menemukan pola pada data serta memprediksi kecenderungan di masa depan. Salah satu metode data mining yang umum digunakan untuk pengelompokan data adalah K-Means Clustering, yaitu metode clustering non-hierarki yang mempartisi data ke dalam satu atau lebih cluster sehingga data dengan karakteristik yang mirip (high intra-class similarity) berada pada cluster yang sama, sedangkan data dengan karakteristik berbeda (low inter-class similarity) berada pada cluster yang berbeda.

Algoritma K-Means dipilih karena memiliki ketelitian yang cukup tinggi terhadap ukuran objek sehingga relatif terukur dan efisien untuk mengolah objek dalam jumlah besar, serta tidak terpengaruh oleh urutan objek. Permasalahan yang dihadapi PT. Sandya

Lestari, sebagai distributor produk NDT, adalah terjadinya penumpukan stok barang yang belum terjual sehingga menjadi pertimbangan penting dalam penyusunan strategi pemasaran dan penjualan yang efektif.

Beberapa penelitian terdahulu memperlihatkan penerapan K-Means Clustering pada berbagai konteks. Mardalius (2018) menunjukkan bahwa hasil pengelompokan aksesoris membantu pemilik toko mengidentifikasi produk paling laku, laku, dan kurang laku. Ghofar dan Kurniawan (2018) menerapkan K-Means untuk mengklasifikasikan pelanggan menjadi kelompok pelanggan potensial, normal, dan tidak potensial. Sabrina Aulia Rahmah (2020) memperoleh tiga kelompok penjualan pestisida dengan proporsi item sangat laku, laku, dan tidak laku yang serupa polanya dengan temuan pada penelitian ini. Wiga Maulana Baihaqi, dkk. (2019) membandingkan K-Means dan K-Medoids dengan bantuan RapidMiner untuk menentukan strategi pemasaran produk. Penelitian-penelitian tersebut menegaskan bahwa K-Means Clustering yang dibantu perangkat lunak seperti RapidMiner efektif digunakan untuk pengelompokan data penjualan guna mendukung keputusan bisnis.

Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis data penjualan barang NDT pada PT. Sandya Lestari menggunakan metode K-Means Clustering, baik melalui perhitungan manual maupun dengan bantuan perangkat lunak RapidMiner Studio, untuk mengelompokkan barang berdasarkan tingkat penjualannya sehingga dapat mendukung penyusunan strategi pemasaran dan pengelolaan stok yang lebih efektif.

LANDASAN TEORI

Data Mining

Data mining adalah proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang sebelumnya tidak disadari keberadaannya, dengan memanfaatkan satu atau lebih teknik pembelajaran komputer (machine learning) untuk mengekstraksi pengetahuan secara otomatis.

Metode K-Means Clustering

K-Means Clustering merupakan salah satu algoritma unsupervised machine learning yang paling sederhana dan populer untuk mempartisi data ke dalam k cluster berdasarkan kedekatan jarak antar-data dengan pusat cluster (centroid). Tahapan umum algoritma K-Means adalah: (1) menentukan jumlah cluster k ; (2) menetapkan centroid awal, umumnya secara acak; (3) menghitung jarak setiap data ke setiap centroid dan menempatkan data pada cluster terdekat; (4) menghitung ulang nilai centroid sebagai rata-rata

anggota tiap cluster; dan (5) mengulangi langkah (3)–(4) hingga posisi centroid tidak lagi berubah.

Jarak antar-data dan centroid dihitung menggunakan Euclidean Distance, yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{[(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2]} \quad (1)$$

dengan $D(i,j)$ adalah jarak data ke- i terhadap pusat cluster ke- j , x_{ki} adalah nilai data ke- i pada atribut ke- k , dan x_{kj} adalah nilai titik pusat cluster ke- j pada atribut ke- k .

RapidMiner

RapidMiner merupakan perangkat lunak open source untuk analisis data mining, text mining, dan analisis prediktif, yang menggunakan berbagai teknik deskriptif dan prediktif untuk membantu pengguna mengambil keputusan. RapidMiner ditulis dengan bahasa Java sehingga dapat berjalan di berbagai sistem operasi, memiliki representasi proses berbasis operator tree, serta mendukung ratusan operator untuk input, output, pra-pemrosesan, pemodelan, dan visualisasi data, termasuk operator Clustering K-Means dan Cluster Distance Performance untuk validasi hasil clustering.

Barang NDT dan Distributor

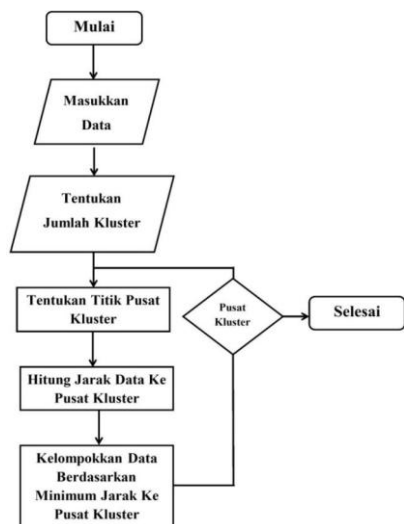
Non-Destructive Testing (NDT) adalah teknik pengujian material tanpa merusak benda uji, mencakup antara lain Liquid Penetrant Testing dan Magnetic Particle Testing (Magnetic Particle Inspection) yang digunakan untuk mendeteksi diskontinuitas permukaan maupun bawah permukaan pada material logam. Distributor adalah pihak yang membeli produk langsung dari produsen dalam jumlah besar untuk disalurkan kepada retailer atau konsumen akhir tanpa memodifikasi produk tersebut.

METODOLOGI

Penelitian ini menggunakan data sekunder berupa data penjualan barang NDT pada PT. Sandya Lestari, yang berlokasi di Jl. Karang Tengah Raya 1 No. 44C, Cilandak. Data dikumpulkan melalui observasi langsung ke lokasi penelitian serta studi literatur terhadap buku, jurnal, dan artikel terkait metode K-Means Clustering. Data yang diolah berjumlah 28 jenis barang dengan tiga atribut, yaitu Stok Awal (SA), Stok Terjual (ST), dan Stok Akhir (SAK), yang selanjutnya dikelompokkan ke dalam 3 cluster.

Tahapan penelitian meliputi: (1) pengumpulan data, (2) penentuan variabel dan jumlah cluster ($k=3$), (3) perhitungan clustering secara manual menggunakan jarak Euclidean dan secara otomatis menggunakan RapidMiner Studio, serta (4) pengujian dan validasi hasil menggunakan Cluster Distance Performance. Centroid awal ditentukan secara acak

dari data, yaitu $C0 = (167, 39, 128)$, $C1 = (1.483, 262, 1.221)$, dan $C2 = (3.860, 210, 3.650)$ berdasarkan atribut (SA, ST, SAK). Proses algoritma K-Means yang digunakan digambarkan pada Gambar 1.



Gambar 1. Flowchart algoritma K-Means

Pengolahan data dengan RapidMiner Studio dilakukan dengan mengimpor data dalam format Microsoft Excel melalui operator Read Excel, menghubungkannya dengan operator Clustering (K-Means) untuk menentukan jumlah cluster dan iterasi maksimal, kemudian menambahkan operator Performance (Cluster Distance Performance) untuk mengevaluasi kinerja hasil pengelompokan berdasarkan rata-rata jarak antara centroid dan seluruh anggota cluster (average within centroid distance). Penelitian dilakukan menggunakan komputer dengan prosesor Intel Core i3 2.3 GHz, RAM 4 GB, VGA 2 GB, dan hard disk 500 GB, dengan sistem operasi Windows 7 64-bit, Microsoft Office, dan RapidMiner Studio.

HASIL DAN PEMBAHASAN

Perhitungan Clustering Manual

Perhitungan jarak setiap data terhadap ketiga centroid awal menggunakan Persamaan (1) menghasilkan penempatan awal 28 item barang ke dalam tiga cluster berdasarkan jarak terdekat. Proses perhitungan diulang dengan memperbarui nilai centroid sebagai rata-rata anggota tiap cluster pada setiap iterasi. Hasil menunjukkan bahwa posisi anggota cluster tidak lagi berubah pada iterasi ke-3, sehingga proses clustering dihentikan pada iterasi tersebut.

Hasil Akhir Pengelompokan

Hasil akhir clustering, baik dari perhitungan manual maupun pengolahan menggunakan RapidMiner Studio, menunjukkan kesesuaian sebagaimana dirangkum pada Tabel 1.

Cluster	Jumlah Item	Kategori	Centroid (SA, ST, SAK)
C0	25	Kurang laku	263,84 ; 54,88 ; 208,96
C1	2	Laku	1.963,50 ; 209 ; 1.754,50
C2	1	Paling laku	3.860 ; 210 ; 3.650

Tabel 1. Ringkasan hasil clustering K-Means

Cluster 2 (C2) hanya beranggotakan satu item, yaitu barang dengan kode UD-ST, dengan Stok Awal, Stok Terjual, dan Stok Akhir tertinggi di antara seluruh item, sehingga dikategorikan sebagai barang paling laku. Cluster 1 (C1) beranggotakan dua item, yaitu UR-ST dan UP-ST, yang dikategorikan sebagai barang laku. Adapun Cluster 0 (C0) beranggotakan 25 item dengan nilai SA, ST, dan SAK relatif rendah, yang dikategorikan sebagai barang kurang laku. Besarnya jumlah anggota C0 dibandingkan C1 dan C2 mengindikasikan bahwa sebagian besar item barang NDT pada PT. Sandya Lestari tergolong kurang laku, sehingga berkontribusi terhadap penumpukan stok barang yang belum terjual.

Validasi Hasil Clustering

Validasi hasil clustering pada RapidMiner Studio menggunakan operator Cluster Distance Performance dengan kriteria average within centroid distance, yaitu rata-rata jarak antara centroid dan seluruh anggota clusternya. Hasil pengujian menunjukkan nilai average within centroid distance sebesar 21.073,928. Nilai ini menggambarkan tingkat kekompakan anggota di dalam masing-masing cluster terhadap centroidnya; semakin kecil nilai tersebut relatif terhadap sebaran data, semakin baik kualitas pengelompokan yang dihasilkan. Visualisasi hasil pengelompokan dalam bentuk diagram pie ditunjukkan pada Gambar 2.



Gambar 2. Proporsi anggota cluster hasil pengolahan RapidMiner Studio

Hasil ini sejalan dengan temuan Sabrina Aulia Rahmah (2020) yang juga memperoleh proporsi anggota cluster tidak merata, dengan jumlah item pada kelompok kurang laku yang jauh lebih besar dibandingkan kelompok laku dan sangat laku. Konsistennya hasil perhitungan manual dengan pengolahan RapidMiner Studio pada penelitian ini menunjukkan bahwa metode K-Means Clustering

dapat diandalkan untuk mengelompokkan data penjualan barang NDT pada PT. Sandya Lestari. Informasi pengelompokan ini dapat menjadi dasar bagi perusahaan dalam menyusun strategi pemasaran yang lebih terfokus, misalnya dengan memberikan promosi atau penawaran khusus pada item-item yang tergolong pada Cluster 0, serta memastikan ketersediaan stok yang memadai untuk item pada Cluster 1 dan Cluster 2 yang permintaannya tinggi.

KESIMPULAN

Berdasarkan hasil analisis, penerapan metode K-Means Clustering pada data penjualan barang NDT PT. Sandya Lestari, baik melalui perhitungan manual maupun menggunakan perangkat lunak RapidMiner Studio, menghasilkan pengelompokan yang konsisten ke dalam 3 cluster dengan posisi centroid konvergen pada iterasi ke-3. Cluster 0 (C0) beranggotakan 25 item barang kurang laku, Cluster 1 (C1) beranggotakan 2 item barang laku, dan Cluster 2 (C2) beranggotakan 1 item barang paling laku, dengan nilai validitas Cluster Distance Performance sebesar 21.073,928. Besarnya jumlah item pada kategori kurang laku menunjukkan bahwa penumpukan stok barang yang belum terjual merupakan permasalahan nyata pada perusahaan, sehingga hasil pengelompokan ini menjadi pertimbangan penting dalam penyusunan strategi pemasaran dan penjualan yang lebih efektif serta pengelolaan persediaan barang yang lebih optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Rektor Universitas Pamulang dan Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Pamulang atas pendanaan dan fasilitasi penelitian ini melalui Kegiatan Penelitian dan Pengabdian kepada Masyarakat Tahun Akademik 2025/2026 (No. Kontrak: 0559/D5/SPKP/LPPM/UNPAM/X/2025), serta kepada seluruh pihak di PT. Sandya Lestari yang telah memberikan izin, data, dan bantuan selama proses penelitian berlangsung.

DAFTAR PUSTAKA

Agus Perdana Windarto. (2017). Penerapan data mining pada ekspor buah-buahan menurut negara tujuan menggunakan K-Means Clustering. *Techno.COM*, 16(4), 348–357.

Deni Triyansyah, & Devi Fitrihanah. (2018). Analisis data mining menggunakan algoritma K-Means Clustering untuk menentukan strategi marketing studi kasus: Hoyweapstore. *IncomTech, Jurnal Telekomunikasi dan Komputer*, 8(3).

Elly Muningsih, & Sri Kiswati. (2015). Penerapan metode K-Means untuk clustering produk online shop dalam penentuan stok barang. *Jurnal Bianglala Informatika*,

Fintri Indriyani, & Eni Irfiani. (2019). Clustering data penjualan pada toko perlengkapan outdoor menggunakan metode K-Means. *Jurnal Informatika*, 7(2).

Ismi Azhami, & Rahmi Fauziah. (2020). Penerapan RapidMiner pada data mining klastering (kasus: distribusi persentase rumah tangga menurut kabupaten/kota dan bahan bakar untuk memasak). *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 1(2), 52–58.

Mardalius. (2018). Pemanfaatan RapidMiner Studio 8.2 untuk pengelompokan data penjualan aksesoris menggunakan algoritma K-Means. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 4(2), 123–132.

Muhammad Abdul Ghofar, & Yogiek Indra Kurniawan. (2018). Aplikasi pengelompokan pelanggan pada UMS Store menggunakan algoritma K-Means. *Jurnal Teknologi & Manajemen Informatika*, 4(1).

Muhammad Iqbal. (2019). Klasterisasi data jamaah umroh pada Auliya Tour & Travel menggunakan metode K-Means Clustering. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 5(2), 97–104.

Nengsi Anggraini, Jasmir, & Pareza Alam Jusia. (2018). Penerapan metode K-Means Clustering untuk menentukan persediaan stok barang pada Toko Pensmart Jambi. *Jurnal Manajemen Sistem Informasi*.

Pareza Alam Jusia, Fadhel Muhammad Irfan, & Kurniabudi. (2019). Clustering data untuk rekomendasi penentuan jurusan perguruan tinggi menggunakan metode K-Means. *Jurnal IKRA-ITH Informatika*, 3(3).

Riyani Wulan Sari, Anjar Wanto, & Agus Perdana Windarto. (2018). Implementasi RapidMiner dengan metode K-Means (studi kasus: imunisasi campak pada balita berdasarkan provinsi). *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2(1).

Sabrina Aulia Rahmah. (2020). Klasterisasi pola penjualan pestisida menggunakan metode K-Means Clustering (studi kasus di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Journal of Information Technology Research*, 1(1).

Siska Haryati, Aji Sudarsono, & Eko Suryana. (2015). Implementasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma C4.5 (studi kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2).

Sri Tria Siska. (2016). Analisa dan penerapan data mining untuk menentukan kubikasi air terjual berdasarkan pengelompokan pelanggan menggunakan algoritma K-Means Clustering. *Jurnal Teknologi Informasi & Pendidikan*, 9(1).

Sri Wahyuni, Kana Saputra S, & Mochammad Iswan Perangin-Angin. (2017). Implementasi RapidMiner dalam menganalisa data mahasiswa drop out. *Jurnal Teknik dan Informatika*, 10(2).

Supriady, Muhammad Ruslan Maulani, & Jonathan Eka Ratmoko. (2018). Penerapan algoritma K-Means pada aplikasi e-commerce untuk mengetahui minat customer di Toko X. *Jurnal Teknik Informatika*, 10(3).

Wiga Maulana Baihaqi, Kuat Indartono, & Syifaal Banat. (2019). Penerapan teknik clustering sebagai strategi pemasaran pada penjualan buku di Tokopedia dan Shopee. *Jurnal Ilmiah Teknologi Informasi*, 21(2).

Wirta Agustin, & Yulya Muharmi. (2020). Penerapan data mining menggunakan algoritma Apriori untuk menentukan pola penyebab gelandangan dan pengemis. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 7(2), 229–234.