Prosiding Seminar Nasional Indonesia
Vol. 4 No. 1 2026, hal., 150-165
https://sociohum.net/index.php/PROSIDINGNASIOANAL
e-ISSN: 3026-5169

KLASTERISASI POLA PERMINTAAN PADA UMKM MENGGUNAKAN DATA TRANSFORMATION DAN K-MEANS CLUSTERING

Lailaturahma Maulidah
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Correspondence author email: 22082010096@student.upnjatim.ac.id

Achmad Mukhlis
Universitas Pembangunan Nasional "Veteran" Jawa Timur
22082010097@student.upnjatim.ac.id

Abstract

The consumer electronics retail industry is characterized by short product life cycles and intermittent demand patterns at the stock keeping unit (SKU) level. This study performs demand pattern segmentation of SKUs at an electronics retail store using the Average Demand Interval (ADI) and Squared Coefficient of Variation (CV^2) features extracted from three years of transaction data. Both features were transformed using log1p and Z-score standardization to address the skewed data distribution, and then clustered using the K-Means algorithm with candidate values of K ranging from 2 to 4. The optimal number of clusters was selected based on the Silhouette Score and Davies-Bouldin Index (DBI), while segment labeling was performed using empirical median thresholds derived from the research data, rather than the standard Syntetos-Boylan-Croston cut-off values. Results on 1,342 valid SKUs show that K=2 is the optimal number of clusters (Silhouette Score 0.6380; DBI 0.7034), which were interpreted as the "routine-stable demand" segment (1,055 SKUs; 78.61%) and the "routine but fluctuating" segment (287 SKUs; 21.39%), with the separation being more dominated by ADI values than by CV^2. These findings indicate that the demand structure is more accurately represented by two transaction-frequency-based groups, and can serve as a basis for inventory management policies tailored to the characteristics of each segment.

Keywords: demand segmentation, clustering, K-Means, Average Demand Interval, Squared Coefficient of Variation, data transformation, electronics retail.

Abstrak

Industri ritel elektronik dicirikan oleh siklus hidup produk pendek dan pola permintaan yang tidak kontinu pada tingkat *stock keeping unit* (SKU). Penelitian ini melakukan segmentasi pola permintaan SKU pada sebuah toko ritel elektronik menggunakan fitur *Average Demand Interval* (ADI) dan *Squared Coefficient of Variation* (CV^2) yang diekstraksi dari data transaksi selama tiga tahun. Kedua fitur ditransformasi menggunakan log1p dan standardisasi Z-score untuk mengatasi distribusi data yang menceng, kemudian dikelompokkan menggunakan algoritma K-Means dengan kandidat K=2 hingga K=4. Pemilihan jumlah klaster optimal

berdasarkan Silhouette Score dan Davies-Bouldin Index (DBI), sedangkan pelabelan segmen menggunakan ambang median empiris data penelitian, bukan cut-off baku Syntetos-Boylan-Croston. Hasil terhadap 1.342 SKU valid menunjukkan K=2 sebagai jumlah klaster optimal (Silhouette Score 0,6380; DBI 0,7034), yang diinterpretasikan sebagai segmen "Permintaan rutin-stabil" (1.055 SKU; 78,61%) dan "Rutin tapi fluktuatif" (287 SKU; 21,39%), dengan pemisahan yang lebih didominasi nilai ADI dibandingkan CV^2. Temuan ini menunjukkan struktur permintaan lebih tepat direpresentasikan oleh dua kelompok berbasis frekuensi transaksi, dan dapat menjadi dasar kebijakan pengelolaan persediaan sesuai karakteristik tiap segmen.

Kata Kunci : segmentasi permintaan, clustering, K-Means, Average Demand Interval, Squared Coefficient of Variation, transformasi data, ritel elektronik.

PENDAHULUAN

Industri ritel beroperasi dalam lingkungan bisnis yang biasanya beroperasi dengan siklus hidup produk yang relatif singkat, tingginya variasi merek dan tipe produk, serta pola pembelian konsumen yang cenderung berfluktuasi. Karakteristik tersebut menyebabkan sebagian besar *stock keeping unit* (SKU) memiliki pola penjualan yang tidak berlangsung secara kontinu, yang ditunjukkan oleh adanya periode tanpa permintaan (*zero demand*) yang berselang-seling dengan periode terjadinya transaksi dalam jumlah yang bervariasi. Dalam literatur manajemen persediaan, kondisi tersebut dikenal sebagai *intermittent demand* atau permintaan jarang, yaitu pola permintaan yang dicirikan oleh interval antar-permintaan yang tidak konstan dan sering kali disertai periode tanpa permintaan.

Kerangka kerja yang paling banyak dirujuk untuk mengklasifikasikan pola permintaan semacam ini adalah skema kategorisasi berbasis *Average Demand Interval* (ADI) dan *Squared Coefficient of Variation* (CV^2) yang diperkenalkan oleh Syntetos, Boylan, dan Croston, dengan threshold ADI = 1,32 dan CV^2 = 0,49 untuk membedakan permintaan menjadi *smooth*, *erratic*, *intermittent*, dan *lumpy (Syntetos et al., 2005)*. Skema ini telah digunakan secara luas dalam berbagai penelitian di sektor manufaktur dan suku cadang, termasuk pada industri elektronik profesional yang melibatkan lebih dari seribu SKU. Pada salah satu penelitian, telah dilakukan evaluasi performa berbagai metode peramalan (SMA13, SES, Croston, dan Syntetos-Boylan Approximation) sekaligus mengusulkan pendekatan simulasi dua tahap untuk memodelkan distribusi permintaan *lumpy* menggunakan kombinasi distribusi *uniform* dan *negative binomial (Solis et al., n.d.)*. Hasil penelitian tersebut menunjukkan bahwa tingginya akurasi statistik suatu metode peramalan tidak selalu berbanding lurus dengan efektivitas pengendalian persediaan. Oleh karena itu, pemahaman yang memadai terhadap karakteristik pola permintaan pada tingkat SKU menjadi prasyarat penting dalam penyusunan kebijakan pengelolaan persediaan.

Perkembangan penelitian terkini juga menunjukkan meningkatnya perhatian terhadap pemanfaatan *machine learning* (ML) dalam analisis permintaan intermiten. Tinjauan sistematis yang dilakukan oleh (Giannopoulos et al., 2025) menunjukkan bahwa penelitian pada bidang ini masih didominasi oleh dua kelompok objek, yaitu suku cadang

(*spare parts*) dan SKU ritel, dengan fokus utama pada penerapan model berbasis jaringan saraf tiruan, algoritma berbasis pohon keputusan (*tree-based*), serta pendekatan hibrida yang mengombinasikan metode peramalan konvensional dengan ML. Meskipun demikian, kajian tersebut juga mengidentifikasi sejumlah kesenjangan penelitian. Salah satunya adalah masih terbatasnya pemanfaatan teknik segmentasi berbasis *clustering* untuk mengelompokkan SKU berdasarkan karakteristik statistik permintaannya sebelum dilakukan pemodelan lanjutan. Selain itu, penelitian yang bersifat *product-agnostic* di luar domain suku cadang dan produk ritel kebutuhan sehari-hari juga masih relatif sedikit. Padahal, produk elektronik konsumen—baik berupa gawai maupun aksesori pendukungnya—memiliki karakteristik yang berpotensi menghasilkan pola permintaan intermiten yang sama kompleksnya, seperti siklus hidup produk yang pendek, tingginya variasi tipe atau IMEI, serta fluktuasi harga jual. Hingga saat ini, karakteristik tersebut masih belum banyak dieksplorasi sebagai objek penelitian dalam studi mengenai permintaan intermiten.

Dalam penerapannya, penelitian (Purnamasari et al., 2023) pada usaha kerajinan kulit skala kecil-menengah di Yogyakarta menunjukkan bahwa algoritma K-Means mampu mengelompokkan produk ke dalam kategori strategis, seperti *prioritized product*, *popular product*, dan *exclusive collection*. Dengan penentuan jumlah klaster menggunakan metode *elbow*, hasil segmentasi tersebut memberikan informasi yang dapat dimanfaatkan sebagai dasar penyusunan kebijakan penyediaan stok. Temuan ini menunjukkan bahwa *clustering* merupakan pendekatan yang relevan untuk melakukan segmentasi produk berdasarkan data transaksi historis, bahkan ketika data yang tersedia masih terbatas pada atribut transaksional tanpa memerlukan rekayasa fitur yang kompleks. Meskipun demikian, kualitas hasil clustering tidak hanya ditentukan oleh algoritma yang digunakan, tetapi juga oleh representasi data sebagai masukan. Hal ini didukung oleh temuan (Sana et al., 2022) dalam penelitian mengenai prediksi *customer churn* pada industri telekomunikasi, yang menunjukkan bahwa penerapan metode transformasi data—termasuk di antaranya transformasi logaritma dan standardisasi Z-score—memberikan pengaruh yang signifikan terhadap performa model dibandingkan data mentah, sebagaimana dibuktikan melalui uji Friedman dan uji lanjut post hoc Holm. Temuan tersebut mengindikasikan bahwa pemilihan metode transformasi data berperan penting dalam meningkatkan kualitas representasi fitur sebelum proses pemodelan dilakukan. Mengingat algoritma K-Means bersifat sensitif terhadap skala dan distribusi data, penelitian ini menerapkan transformasi logaritma (log1p) yang dilanjutkan dengan standardisasi Z-score terhadap fitur ADI dan CV^2 sebelum tahap clustering dilakukan, guna mengatasi karakteristik distribusi kedua fitur yang cenderung menceng (*skewed*) pada data transaksi ritel. Namun demikian, penerapan transformasi data pada segmentasi pola permintaan berbasis ADI-CV^2, khususnya pada data transaksi ritel elektronik konsumen, masih relatif jarang dibahas dalam literatur.

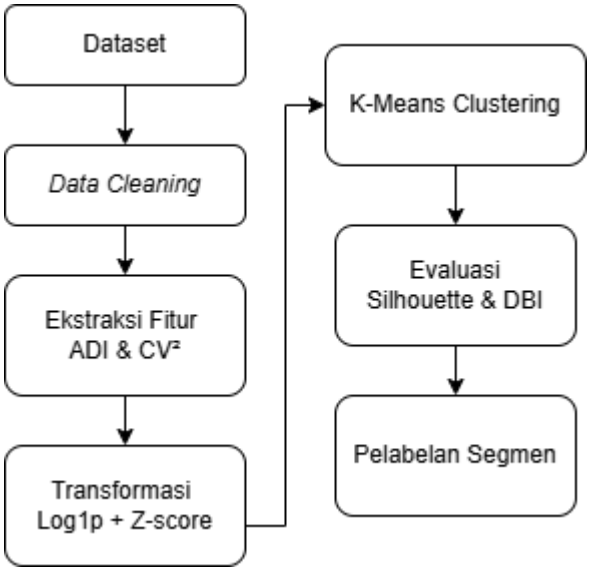
Berdasarkan kajian tersebut, terdapat beberapa kesenjangan penelitian yang masih perlu dikaji lebih lanjut. Pertama, penerapan skema kategorisasi ADI-CV^2 pada data transaksi ritel elektronik konsumen masih terbatas, sementara sebagian besar penelitian sebelumnya berfokus pada industri dengan skala permintaan yang cukup besar. Padahal, karakteristik produk elektronik konsumen, seperti siklus hidup produk

yang pendek, tingginya variasi tipe, serta fluktuasi permintaan, berpotensi menghasilkan pola permintaan yang berbeda. Kedua, pengaruh berbagai metode transformasi data terhadap kualitas hasil *clustering* pola permintaan belum banyak dievaluasi secara sistematis menggunakan metrik validasi internal, seperti *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Ketiga, penentuan jumlah klaster pada penelitian-penelitian sebelumnya umumnya hanya mengandalkan satu metode evaluasi, misalnya *elbow*, sehingga belum banyak penelitian yang membandingkan beberapa kandidat jumlah klaster menggunakan lebih dari satu metrik validasi.

Untuk menjawab kesenjangan tersebut, penelitian ini mengusulkan pendekatan segmentasi pola permintaan berbasis fitur *Average Demand Interval* (ADI) dan *Squared Coefficient of Variation* (CV^2) yang diekstraksi dari data transaksi historis selama tiga tahun pada Toko XYZ, sebuah toko ritel elektronik yang menjual *smartphone*, tablet, perangkat audio, serta berbagai aksesori pendukung. Data transaksi yang mencakup atribut merek, nama barang, tipe barang, no imei, tgl terjual, nama pembeli, beli dari, dan harga jual terlebih dahulu diagregasi menjadi deret waktu penjualan setiap SKU. Selanjutnya, nilai ADI dan CV^2 dihitung sebagai representasi karakteristik pola permintaan sesuai kerangka kategorisasi Syntetos-Boylan-Croston yang telah digunakan oleh (Solis et al., n.d.) dan dirangkum kembali oleh (Giannopoulos et al., 2025).

Tahap berikutnya menerapkan beberapa metode transformasi data terhadap fitur ADI dan CV^2 sebelum proses *clustering* menggunakan algoritma K-Means. Jumlah klaster dibatasi pada rentang dua hingga empat dengan mempertimbangkan jumlah kategori pola permintaan pada skema Syntetos-Boylan-Croston. Selanjutnya, kombinasi metode transformasi data dan jumlah klaster terbaik dipilih berdasarkan dua metrik validasi internal, yaitu *Silhouette Score* dan *Davies-Bouldin Index*, sehingga proses evaluasi tidak bergantung pada satu indikator saja. Dari pendekatan yang digunakan, hasil penelitian ini diharapkan dapat memberikan dasar yang lebih kuat dalam memahami karakteristik pola permintaan produk elektronik konsumen. Selain itu, temuan yang diperoleh juga berpotensi mendukung pengambilan keputusan dalam pengelolaan persediaan melalui pemanfaatan hasil segmentasi SKU yang lebih sesuai dengan kondisi permintaan aktual.

METODE PENELITIAN



Gambar 1. Alur Penelitian

Penelitian ini dilaksanakan melalui beberapa tahap yang akan dirincikan lebih lanjut pada poin-poin berikut:

1. Pengumpulan data transaksi ritel elektronik selama tiga tahun
2. Agregasi data menjadi deret waktu penjualan per SKU
3. Ekstraksi fitur ADI dan CV^2 berdasarkan skema Syntetos-Boylan-Croston
4. Transformasi data menggunakan log1p dan Z-score
5. *Clustering* K-Means dengan kandidat jumlah klaster K=2 hingga K=4
6. Evaluasi internal menggunakan *Silhouette Score* dan *Davies-Bouldin Index* (DBI)
7. Pelabelan segmen berdasarkan ambang median empiris ADI dan CV^2
8. Penyusunan profil segmen sebagai dasar rekomendasi kebijakan persediaan.

Sumber dan Periode Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari catatan transaksi penjualan historis pada Toko XYZ, sebuah usaha ritel yang bergerak di bidang penjualan produk elektronik konsumen, meliputi *smartphone*, tablet, perangkat audio, serta berbagai aksesori pendukung. Data dikumpulkan dalam bentuk *excel* rekapitulasi transaksi penjualan yang mencakup rentang waktu selama tiga tahun. Pemilihan rentang waktu tiga tahun dilakukan dengan pertimbangan bahwa perhitungan interval antar-permintaan (*Average Demand Interval*) memerlukan riwayat transaksi yang cukup panjang agar pola kemunculan permintaan tiap SKU, khususnya untuk produk dengan siklus hidup pendek dan variasi tipe yang tinggi yang dapat terepresentasi secara memadai dan tidak bias akibat periode observasi yang terlalu singkat.

Data yang digunakan adalah data transaksi dengan satuan pengamatan berupa baris transaksi penjualan individual (baris per transaksi). Variabel yang tercakup dalam

data mentah meliputi merek, nama barang, tipe barang, nomor IMEI, dan tanggal terjual. Dari variabel tersebut, tanggal transaksi dan identitas SKU (kombinasi merek, nama barang, dan tipe barang) merupakan variabel utama yang digunakan untuk membentuk deret waktu penjualan setiap SKU, sedangkan nomor IMEI digunakan untuk pengenalan unik SKU guna menghindari duplikasi data per SKU-nya. Dari deret waktu tersebut selanjutnya diturunkan dua variabel turunan (*derived features*) yang menjadi fokus analisis, yaitu *Average Demand Interval* (ADI) yang merepresentasikan rata-rata jarak waktu antar kemunculan transaksi suatu SKU, dan *Squared Coefficient of Variation* (CV^2) yang merepresentasikan tingkat fluktuasi kuantitas permintaan pada saat transaksi terjadi. ADI merepresentasikan rata-rata jarak waktu antar kemunculan transaksi suatu SKU, dihitung dengan formula:

$$ADI = \frac{Total\ Hari\ Periode}{Jumlah\ Hari\ Aktif\ Terjual}$$

Sedangkan CV^2 merepresentasikan tingkat variasi kuantitas permintaan pada saat transaksi terjadi, dihitung dengan formula:

$$CV^2 = \left(\frac{\sigma_{qty}}{\mu_{qty}}\right)^2$$

Semakin besar nilai ADI, semakin jarang suatu SKU mengalami transaksi. Semakin besar nilai CV^2, semakin tidak konsisten kuantitas yang terjual pada setiap transaksi SKU tersebut. Kedua variabel ini mengacu pada kerangka kategorisasi pola permintaan yang diperkenalkan oleh (Syntetos et al., 2005).

NO	MEREK	NAMA BARANG	TIPE BARANG	NO IMEI	TGL TERJUAL	NAMA PEMBELI	BELI DARI	HARGA JUAL
1	SAMSUNG	HANDPHONE	S20 FE 8/128GB LAVENDER	350704160373976	2023-01-01 00:00:00		TOPED	55642650
2	APPLE	HANDPHONE	IPHONE 11 64GB BLACK	354496527369443	2023-01-01 00:00:00		TOPED	81252650
3	SAMSUNG	HANDPHONE	A04E 3/32GB BLUE	352129771016087	2023-01-01 00:00:00		TOPED	12066250
4	SAMSUNG	HANDPHONE	S21 FE 8/128GB GRAPHITE	352254420125954	2023-01-01 00:00:00		TOPED	77212650
6	APPLE	HANDPHONE	IPHONE 11 64GB BLACK	354496527755922	2023-01-01 00:00:00		SHOPEE	81252650
7	APPLE	HANDPHONE	IPHONE 14 PRO 256GB SILVER	351055827594226	2023-01-01 00:00:00		SHOPEE	239837650
8	SAMSUNG	HANDPHONE	A04E 3/32GB BLACK	352129771386662	2023-01-01 00:00:00		TOPED	12066250
9	SAMSUNG	TAB	TAB S6 LITE 4/128GB BLUE	354241700311162	2023-01-01 00:00:00		COD	54090000

Gambar 2. Data Mentah Transaksi Toko XYZ

Seluruh analisis dilakukan menggunakan bahasa pemrograman Python, dengan memanfaatkan pustaka pandas untuk manipulasi dan agregasi data, numpy untuk transformasi log1p guna mereduksi *skewness* distribusi ADI dan CV2, serta scikit-learn untuk proses *clustering* menggunakan algoritma K-Means, evaluasi jumlah *cluster* optimal melalui *Silhouette Score*, *Davies-Bouldin Index* (DBI) dan normalisasi fitur menggunakan *StandardScaler* (Z-score). Pendekatan ini memastikan metodologi bersifat *reproducible* dan dapat direplikasi secara independen.

Teknik Analisis Data

Analisis data dalam penelitian ini dilakukan melalui beberapa tahapan yang saling berurutan. Tahap pertama adalah *preprocessing* data, yaitu pembersihan dan agregasi

data transaksi mentah menjadi deret waktu penjualan harian per SKU, sekaligus penyaringan SKU yang tidak memiliki riwayat transaksi memadai untuk dihitung nilai ADI dan CV^2-nya. Tahap kedua adalah ekstraksi fitur, yaitu perhitungan nilai ADI dan CV^2 untuk setiap SKU berdasarkan deret waktu yang telah dibentuk, mengikuti formulasi yang digunakan dalam kerangka Syntetos-Boylan-Croston.

Tahap ketiga adalah transformasi data, yang diterapkan untuk mengatasi karakteristik distribusi ADI dan CV^2 yang cenderung menceng (*skewed*) pada data transaksi ritel. Transformasi yang digunakan adalah log1p, yaitu transformasi logaritma natural dengan penambahan konstanta satu ($\log(1+x)$) yang berfungsi menstabilkan varians sekaligus menangani nilai nol pada data CV^2, dilanjutkan dengan standardisasi Z-score agar kedua fitur berada pada skala yang setara sebelum digunakan sebagai masukan algoritma *clustering* yang sensitif terhadap skala data, seperti K-Means. Pemilihan pendekatan ini merujuk pada temuan (Sana et al., 2022)yang menunjukkan bahwa metode transformasi data, termasuk transformasi logaritma dan standardisasi, memberikan pengaruh signifikan terhadap performa model berbasis data numerik.

Tahap keempat adalah proses *clustering* menggunakan algoritma K-Means terhadap fitur ADI dan CV^2 yang telah ditransformasi. Jumlah klaster yang diuji dibatasi pada rentang dua hingga empat klaster, dengan pertimbangan jumlah kategori pola permintaan yang lazim digunakan dalam skema kategorisasi Syntetos-Boylan-Croston, yaitu *smooth*, *erratic*, *intermittent*, dan *lumpy*. Pemilihan jumlah klaster optimal dilakukan berdasarkan dua metrik evaluasi internal secara bersamaan, yaitu *Silhouette Score*, yang mengukur seberapa baik pemisahan antar klaster, dan *Davies-Bouldin Index* (DBI), yang mengukur tingkat kekompakan dan keterpisahan klaster, sehingga keputusan tidak hanya bergantung pada satu indikator evaluasi tunggal.

Tahap terakhir adalah interpretasi dan pelabelan segmen, yaitu pemberian label deskriptif terhadap masing-masing klaster yang terbentuk berdasarkan karakteristik rata-rata dan median nilai ADI serta CV^2 pada tiap klaster, sehingga hasil segmentasi dapat diinterpretasikan secara kontekstual dan relevan dengan kebutuhan pengelolaan persediaan produk elektronik konsumen.

Pelabelan Segmen

Perlu ditegaskan bahwa nilai ambang yang digunakan untuk memberi label segmen pada penelitian ini adalah median empiris dari seluruh SKU dalam *dataset*, bukan nilai *cut-off* baku dari skema Syntetos-Boylan-Croston (ADI=1,32 dan CV^2=0,49). Keputusan ini diambil dengan pertimbangan bahwa nilai *cut-off* tersebut pada dasarnya diturunkan untuk tujuan spesifik, yaitu meminimalkan selisih *mean-squared-error* antara *estimator Croston* dan pendekatan Syntetos-Boylan dalam pemilihan metode peramalan, bukan sebagai batas universal untuk segmentasi permintaan pada skala data apa pun. Selain itu, validitas *cut-off* tetap tersebut juga telah dipertanyakan oleh

(Kostenko & Hyndman, 2006), yang menemukan bahwa batas antar kategori permintaan pada kerangka Croston-SBA sesungguhnya bersifat non-linear, bukan garis lurus sebagaimana diasumsikan dalam skema klasik. Dengan demikian, penggunaan ambang median empiris yang diturunkan langsung dari distribusi data penelitian ini merupakan keputusan metodologis yang disengaja untuk menyesuaikan skema pelabelan dengan karakteristik data transaksi ritel elektronik yang dianalisis, sejalan dengan kecenderungan penelitian terkini yang mengadopsi pendekatan segmentasi permintaan berbasis data (*data-driven*) sebagai alternatif dari *threshold* tetap yang bersifat generik.

HASIL DAN PEMBAHASAN

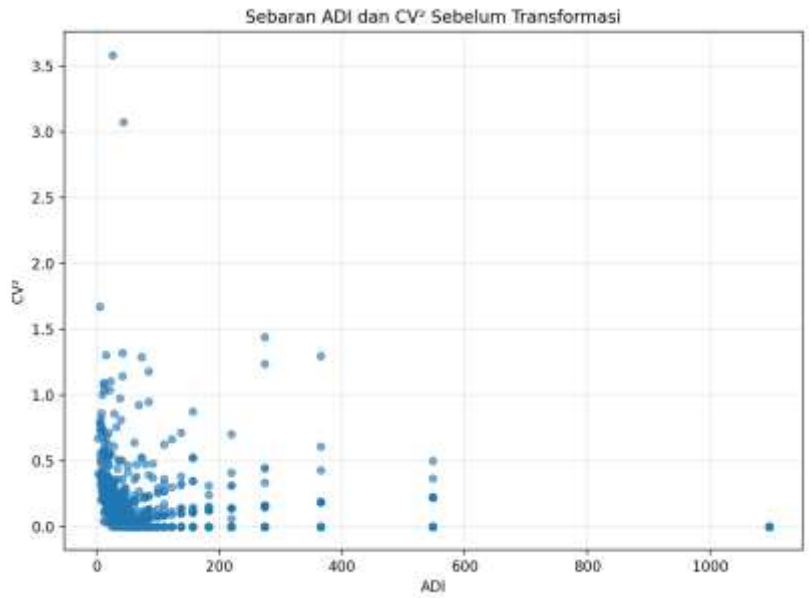
Ekstraksi Fitur Permintaan (ADI dan $CV^{2)}$

Tahap awal analisis dilakukan dengan mengekstraksi dua fitur statistik dari data transaksi bersih setiap SKU, yaitu *Average Demand Interval* (ADI) dan *Squared Coefficient of Variation* (CV^2). ADI merepresentasikan rata-rata jarak waktu antar kemunculan transaksi suatu SKU, sedangkan CV^2 merepresentasikan tingkat variasi kuantitas permintaan pada saat transaksi terjadi. Semakin besar nilai ADI, semakin jarang suatu SKU mengalami transaksi. Semakin besar nilai CV^2, semakin tidak konsisten kuantitas yang terjual pada setiap transaksi SKU tersebut. Proses ekstraksi fitur ini menghasilkan sebanyak 1.342 SKU valid yang memiliki riwayat transaksi memadai untuk dihitung nilai ADI dan CV^2-nya sepanjang periode tiga tahun observasi.

Tabel 1. Statistik Deskriptif ADI dan CV^2 (Nilai Asli, Sebelum Transformasi)

Statistik	ADI	CV^2
Nilai minimum	1,4891	0,0000
Nilai maksimum	1.096,0000	3,5827
Rata-rata (mean)	561,7267	0,0907
Median	548,0000	0,0000
Standar deviasi	452,3809	0,2334
Skewness	0,1603	6,1154

Berdasarkan Tabel 1, terlihat bahwa nilai *skewness* ADI relatif kecil (0,1603), yang mengindikasikan sebaran ADI mendekati simetris meskipun rentangnya cukup lebar (1,49 hingga 1.096 hari). Sebaliknya, nilai *skewness* CV^2 sangat tinggi (6,1154), menandakan distribusi yang sangat menceng ke kanan (*right-skewed*). Kondisi ini terjadi karena median CV^2 bernilai 0 — mengindikasikan bahwa lebih dari separuh SKU dalam dataset ini tidak memiliki variasi kuantitas penjualan sama sekali pada hari-hari aktif transaksinya (standar deviasi kuantitas = 0), sementara sebagian kecil SKU lainnya memiliki CV^2 yang jauh lebih tinggi (hingga 3,5827). Tingginya *skewness* pada CV^2 inilah yang menjadi dasar pertimbangan perlunya transformasi data sebelum fitur ini digunakan sebagai masukan algoritma *clustering* berbasis jarak seperti K-Means.



Gambar 3. Sebaran Nilai ADI dan CV² Seluruh SKU Sebelum Transformasi

Gambar 3 memperlihatkan bahwa sebagian besar titik data menumpuk pada sumbu CV² = 0, sementara nilai ADI tersebar cukup lebar sepanjang sumbu horizontal. Pola sebaran ini menegaskan temuan pada Tabel 1 bahwa distribusi CV² sangat tidak merata dan terkonsentrasi pada satu titik, sehingga apabila fitur mentah ini langsung digunakan untuk *clustering*, algoritma berisiko didominasi oleh perbedaan skala ADI semata dan kurang sensitif terhadap variasi CV² yang sesungguhnya informatif pada sebagian kecil SKU.

Transformasi Data

Untuk mengatasi permasalahan distribusi yang menceng sebagaimana ditunjukkan pada Sub-bab 4.1, diterapkan dua tahap transformasi data secara berurutan terhadap fitur ADI dan CV² sebelum digunakan dalam proses clustering:

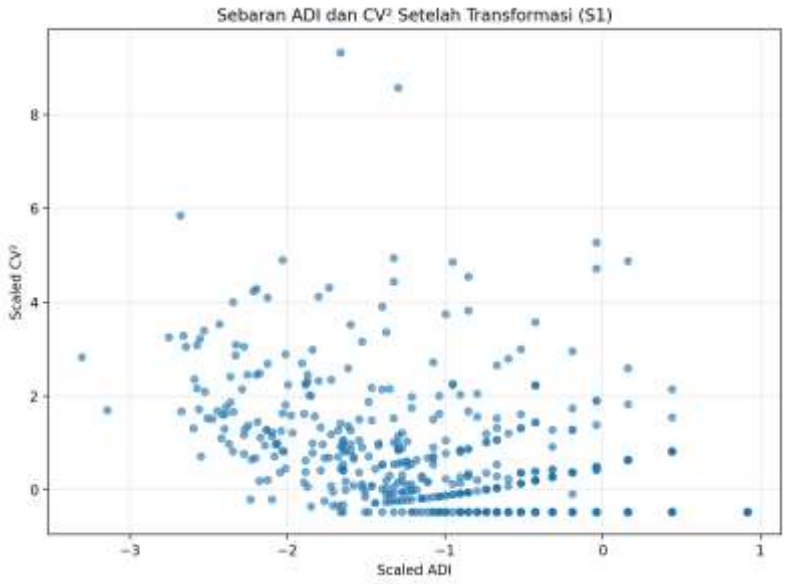
1. Transformasi log1p, yaitu log1p(x) = ln(1 + x), diterapkan untuk mereduksi *skewness* ekstrem terutama pada fitur CV² (*skewness* = 6,1154). Penambahan konstanta 1 pada rumus ini memastikan transformasi tetap aman digunakan meskipun terdapat nilai 0 pada data, yang tidak dapat ditangani oleh transformasi logaritma natural biasa.
2. Standardisasi Z-score, diterapkan setelah log1p untuk menyamakan skala kedua fitur, mengingat algoritma K-Means menghitung jarak *Euclidean* antar titik data sehingga sensitif terhadap perbedaan skala antar fitur.

Tabel 2. Perbandingan Distribusi ADI dan CV² Sebelum dan Sesudah Transformasi

Fitur	Tahap	Mean	Std Dev	Skewness
ADI	Raw	561,7267	452,3809	0,1603

ADI	Setelah log1p	5,6743	1,4413	−0,8710
ADI	Setelah Z-Score	≈ 0,0000	≈ 1,0000	—
CV^2	Raw	0,0907	0,2334	6,1154
CV^2	Setelah log1p	0,0722	0,1555	3,2965
CV^2	Setelah Z-Score	≈ 0,0000	≈ 1,0000	—

Transformasi log1p terbukti efektif dalam mereduksi *skewness* fitur ADI, dari 0,1603 pada data mentah menjadi −0,8710 setelah transformasi, nilai yang jauh lebih mendekati nol dan menunjukkan distribusi yang lebih simetris. Namun, pada fitur CV^2, meskipun log1p berhasil menurunkan *skewness* dari 6,1154 menjadi 3,2965, nilai tersebut masih tergolong tinggi. Hal ini terjadi karena akar permasalahan pada fitur CV^2 bukan sekadar rentang nilai yang lebar, melainkan proporsi SKU dengan nilai $CV^2 = 0$ yang sangat dominan (tercermin dari median $CV^2 = 0$ pada Tabel 2), sehingga transformasi logaritma tidak dapat sepenuhnya menghilangkan konsentrasi nilai pada titik nol tersebut. Meskipun demikian, transformasi log1p yang dilanjutkan dengan standardisasi Z-score tetap diperlukan agar kedua fitur berada pada skala yang setara sebagai masukan algoritma K-Means, dan efeknya terhadap kualitas hasil *clustering* dievaluasi lebih lanjut pada subab selanjutnya.



Gambar 4. Sebaran ADI dan CV^2 Setelah Transformasi Log1p dan Z-score

Dibandingkan dengan Gambar 3, sebaran data pada Gambar 4 menunjukkan distribusi yang lebih merata pada sumbu ADI setelah transformasi, meskipun konsentrasi titik pada nilai CV^2 rendah masih tampak, sejalan dengan temuan skewness CV^2 yang belum sepenuhnya tereduksi pada Tabel 2.

Penentuan Jumlah Cluster Optimal (K)

Algoritma K-Means memerlukan jumlah cluster (K) yang ditentukan di awal. Penelitian ini mengevaluasi kandidat K = 2, 3, dan 4, mengacu pada jumlah kategori pola permintaan pada skema Syntetos-Boylan-Croston yang menggunakan dua metrik validasi internal:

- Silhouette Score, mengukur seberapa mirip suatu titik data dengan cluster miliknya sendiri dibandingkan dengan cluster lain, dengan rentang nilai –1 hingga 1. Semakin tinggi nilainya, semakin baik pemisahan antar cluster.
- Davies-Bouldin Index (DBI), mengukur rasio jarak dalam cluster (*intra-cluster*) terhadap jarak antar cluster (*inter-cluster*), dengan nilai ≥ 0. Semakin rendah nilainya, semakin kompak dan terpisah cluster yang terbentuk.

Pemilihan K terbaik dilakukan dengan memprioritaskan nilai DBI terendah, dengan *Silhouette Score* tertinggi sebagai kriteria penguat apabila terdapat kesamaan nilai DBI antar kandidat K.

Tabel 3. Hasil Evaluasi Metrik Clustering untuk K = 2, 3, dan 4

K	Silhouette Score	Davies-Bouldin Index
2	0,6380	0,7034
3	0,6022	0,7317
4	0,5979	0,7567

Berdasarkan Tabel 3, K = 2 dipilih sebagai jumlah *cluster* optimal karena secara konsisten menghasilkan performa terbaik pada kedua metrik evaluasi sekaligus: DBI terendah (0,7034) dan Silhouette Score tertinggi (0,6380) dibandingkan K = 3 dan K = 4. Pola ini menunjukkan bahwa penambahan jumlah *cluster* dari 2 menjadi 3 atau 4 justru menurunkan kualitas pemisahan *cluster* yang mengindikasikan bahwa struktur alami data ADI-CV^2 pada dataset ini paling baik direpresentasikan oleh dua kelompok, bukan empat kelompok sebagaimana diasumsikan dalam kerangka kategorisasi Syntetos-Boylan-Croston klasik.

Hasil Clustering K-Means (K = 2)

Setelah K = 2 ditetapkan sebagai jumlah cluster optimal, algoritma K-Means dijalankan pada fitur ADI dan CV^2 yang telah ditransformasi (log1p + Z-score). Proses ini menghasilkan dua cluster dengan distribusi SKU sebagai berikut.

Tabel 4. Distribusi SKU per Cluster

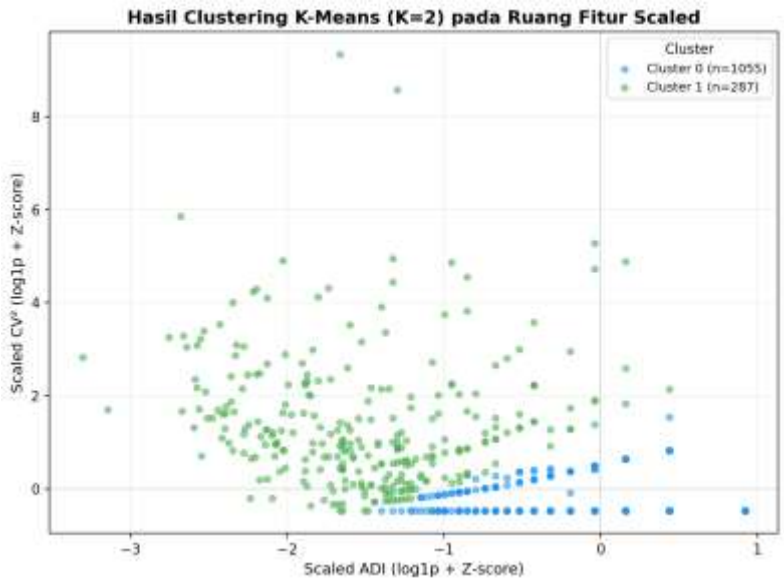
Cluster	Jumlah SKU	Persentase
Cluster 0	1.055	78,61%
Cluster 1	287	21,39%
Total	1.342	100%

Distribusi pada Tabel 4 menunjukkan bahwa *Cluster* 0 mendominasi dataset dengan mencakup 78,61% dari total SKU, sementara *Cluster* 1 mencakup sisanya sebesar 21,39%. Untuk memahami karakteristik masing-masing cluster secara lebih rinci, profil statistik ADI dan CV^2 pada nilai aslinya (belum ditransformasi) ditampilkan pada Tabel 5.

Tabel 5. Profil Statistik Tiap Cluster (Nilai Asli ADI dan CV^2)

Cluster	n SKU	ADI mean	ADI median	ADI min	ADI max	CV^2 mean	CV^2 median	CV^2 min	CV^2 max
Cluster 0	1.055	700,26	548	37,79	1.096	0,0142	0,0000	0,00	0,3673
Cluster 1	287	52,47	32,24	1,49	548	0,3715	0,2678	0,00	3,5827

Tabel 5 menunjukkan karakteristik yang kontras antara kedua *cluster*. *Cluster* 0 memiliki ADI rata-rata jauh lebih tinggi (700,26 hari) dibandingkan *Cluster* 1 (52,47 hari), namun dengan nilai CV^2 rata-rata yang jauh lebih rendah (0,0142 berbanding 0,3715). Dengan kata lain, pemisahan kedua *cluster* ini tampak lebih didominasi oleh sumbu ADI dibandingkan sumbu CV^2, mengingat rentang ADI kedua *cluster* hampir tidak beririsan (*Cluster* 1 maksimum 548,00, tepat pada titik median *Cluster* 0), sedangkan rentang CV^2 keduanya sama-sama dimulai dari 0.



Gambar 5. Hasil Clustering K-Means pada Ruang Fitur Scaled ADI dan CV^2

Gambar 5 memperlihatkan pemisahan geometris kedua cluster pada ruang fitur yang telah ditransformasi, di mana Cluster 0 dan Cluster 1 membentuk dua kelompok titik yang relatif terpisah.

Interpretasi Segmen Permintaan Berbasis Median Empiris

Setelah cluster teknis terbentuk, tahap selanjutnya adalah memberikan label segmen yang bermakna secara bisnis terhadap masing-masing cluster. Berbeda dengan kerangka Syntetos-Boylan-Croston klasik yang menggunakan ambang tetap (ADI = 1,32 dan CV^2 = 0,49), penelitian ini menggunakan ambang median empiris yang diturunkan langsung dari distribusi data SKU pada dataset penelitian ini.

Tabel 6. Batas Definisi Keempat Segmen Permintaan

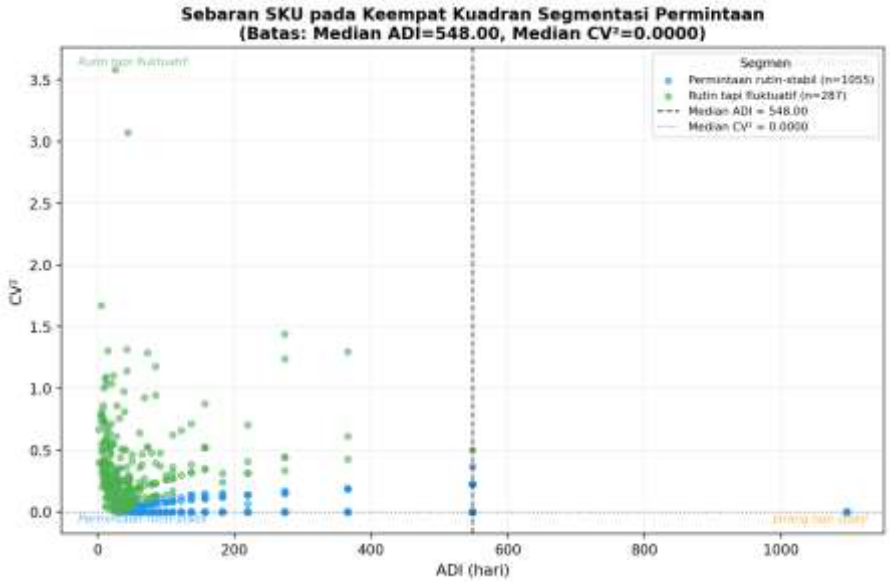
Segmen	Kondisi ADI	Kondisi CV^2	Keterangan
Rutin-stabil	≤ 548	≤ 0	Sering terjual, tidak ada variasi kuantitas
Jarang tapi stabil	> 548	≤ 0	Jarang terjual, tidak ada variasi kuantitas
Rutin tapi fluktuatif	≤ 548	> 0	Sering terjual, ada variasi kuantitas
Jarang dan flutuatif	> 548	> 0	Jarang terjual, ada variasi kuantitas

Namun, ketika kedua *cluster* hasil K-Means dipetakan terhadap keempat kemungkinan segmen tersebut, hanya dua dari empat segmen yang terisi oleh data aktual.

Tabel 7. Pemetaan Cluster ke Segmen Aktual

Cluster	ADI median	CV^2 median	Posisi terhadap Threshold	Segmen Aktual
0	548,00	0,0000	ADI ≤ 548 dan CV^2 ≤ 0	Permintaan rutin-stabil
1	32,24	0,2678	ADI ≤ 548 dan CV^2 > 0	Rutin tapi fluktuatif

Sebagaimana ditunjukkan pada Tabel 7, kedua *cluster* sama-sama memiliki nilai ADI median ≤ 548 (*threshold* penelitian ini), sehingga tidak ada satu pun *cluster* yang jatuh pada kuadran jarang (ADI > 548). Kondisi ini mengindikasikan bahwa secara kolektif, populasi SKU pada dataset penelitian ini didominasi oleh produk dengan interval permintaan yang relatif tidak jauh dari titik tengah distribusinya sendiri dan ini merupakan hal yang wajar terjadi karena ambang yang digunakan adalah median dari populasi yang sama, sehingga secara definisi separuh populasi akan selalu berada di ≤ median.



Gambar 6. Sebaran SKU pada Segmen Permintaan dengan Batas Median ADI dan CV^2

Gambar 6 menampilkan sebaran SKU pada ruang ADI × CV^2 (nilai asli) dengan garis vertikal pada ADI = 548 dan garis horizontal pada CV^2 = 0, mengonfirmasi secara visual bahwa kuadran kanan atas dan kanan bawah (segmen "jarang tapi stabil" dan "jarang dan fluktuatif") tidak terisi oleh SKU mana pun pada dataset ini.

Analisis/Diskusi

Tidak terisinya kuadran jarang bukan disebabkan oleh ketiadaan produk dengan interval permintaan panjang, melainkan konsekuensi langsung dari penggunaan ambang median populasi itu sendiri sebagai batas pemisah: karena median dihitung dari populasi yang sama, secara definisi setengah populasi akan selalu berada pada sisi ADI ≤ median. Namun, hasil clustering K-Means pada penentuan jumlah klaster optimal justru menunjukkan bahwa pemisahan dua *cluster* yang optimal (K=2) terjadi tepat pada sumbu ADI, bukan pada kombinasi ADI-CV^2 sebagaimana diasumsikan kerangka teoritis empat kuadran. Hal ini mengindikasikan bahwa struktur alami data pada dataset ini lebih baik dijelaskan oleh dua kelompok berbasis frekuensi transaksi, dibandingkan empat kelompok berbasis kombinasi frekuensi dan variabilitas kuantitas.

Implikasi median CV^2 = 0 Ambang CV^2 = 0 menyebabkan segmentasi berdasarkan CV^2 menjadi bersifat biner, yaitu SKU yang pernah memiliki variasi kuantitas penjualan (CV^2 > 0) versus SKU yang sama sekali tidak pernah memiliki variasi (CV^2 = 0). Kondisi ini berbeda secara mendasar dari makna CV^2 pada kerangka Syntetos-Boylan-Croston klasik, yang menggunakan ambang CV^2 = 0,49 untuk membedakan tingkat volatilitas kuantitas permintaan pada SKU yang sama-sama sudah memiliki variasi. Perbedaan ini mencerminkan karakteristik data transaksi UMKM yang diteliti, yaitu volume transaksi per hari yang relatif rendah sehingga banyak SKU hanya terjual dalam kuantitas tunggal

(satu unit) pada setiap kejadian transaksi, mengakibatkan standar deviasi kuantitas yang menyebabkan CV^2 bernilai nol.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah diuraikan, dapat disimpulkan bahwa pendekatan segmentasi pola permintaan berbasis fitur ADI dan CV^2 dengan transformasi log1p dan standardisasi Z-score berhasil mengelompokkan 1.342 SKU pada data transaksi ritel elektronik ke dalam dua klaster yang optimal. Pemilihan K=2 sebagai jumlah klaster terbaik, yang didukung oleh Silhouette Score sebesar 0,6380 dan Davies-Bouldin Index sebesar 0,7034, menunjukkan bahwa struktur permintaan pada dataset ini lebih baik direpresentasikan oleh dua kelompok dibandingkan tiga atau empat kelompok sebagaimana diasumsikan dalam kerangka kategorisasi Syntetos-Boylan-Croston klasik.

Kedua klaster yang terbentuk, yaitu "Permintaan rutin-stabil" (78,61% dari total SKU) dan "Rutin tapi fluktuatif" (21,39% dari total SKU), menunjukkan bahwa pemisahan klaster lebih didominasi oleh perbedaan nilai ADI dibandingkan CV^2, mengingat kedua klaster sama-sama memiliki nilai ADI median di bawah atau sama dengan ambang median populasi. Kondisi ini menyebabkan tidak terisinya dua dari empat kemungkinan segmen dalam kerangka teoritis (kategori "jarang"), yang merupakan konsekuensi langsung dari penggunaan ambang median empiris populasi itu sendiri sebagai batas segmentasi, bukan indikasi ketiadaan variasi interval permintaan pada data. Temuan ini turut menegaskan bahwa penggunaan ambang cut-off baku (ADI=1,32 dan CV^2=0,49) yang dirancang untuk konteks pemilihan metode peramalan tidak dapat diterapkan secara langsung pada data dengan skala dan granularitas yang berbeda, sehingga pendekatan ambang median empiris yang diadopsi dalam penelitian ini lebih relevan digunakan pada karakteristik data transaksi ritel elektronik dengan volume transaksi harian yang relatif rendah.

Hasil segmentasi ini dapat dimanfaatkan sebagai dasar penyusunan kebijakan pengelolaan persediaan yang berbeda untuk masing-masing segmen: SKU pada segmen "Permintaan rutin-stabil" dapat dikelola dengan strategi stok minimal tanpa buffer besar karena tidak menunjukkan variasi kuantitas permintaan, sedangkan SKU pada segmen "Rutin tapi fluktuatif" memerlukan buffer stok yang lebih fleksibel mengingat tingkat variasi kuantitas permintaannya yang lebih tinggi.

Penelitian ini memiliki beberapa keterbatasan yang dapat menjadi arah penelitian selanjutnya. Pertama, penggunaan ambang median populasi menyebabkan segmentasi secara struktural selalu menghasilkan pembagian yang seimbang di sekitar titik tengah data, sehingga kategori "jarang" tidak dapat teramati meskipun terdapat variasi interval permintaan yang cukup lebar dalam dataset; penelitian selanjutnya dapat mengeksplorasi metode penentuan ambang alternatif, seperti persentil non-median atau pendekatan berbasis domain, untuk mengakomodasi kemunculan

keempat kategori segmen secara lebih seimbang. Kedua, penelitian ini terbatas pada satu jenis transformasi data (log1p dan Z-score) tanpa membandingkan dengan metode transformasi lain, sehingga belum diketahui secara pasti sejauh mana pilihan metode transformasi memengaruhi kualitas hasil clustering pada data serupa. Ketiga, penelitian ini terbatas pada satu toko ritel elektronik sehingga generalisasi hasil terhadap toko ritel elektronik lain dengan karakteristik produk dan volume transaksi yang berbeda perlu dikaji lebih lanjut pada penelitian mendatang.

DAFTAR PUSTAKA

Giannopoulos, P. G., Dasaklis, T. K., Tsantilis, I., & Patsakis, C. (2025). Machine learning algorithms in intermittent demand forecasting: a review. In *International Journal of Production Research*. Taylor and Francis Ltd. https://doi.org/10.1080/00207543.2025.2578701

Kostenko, A. V., & Hyndman, R. J. (2006). A note on the categorization of demand patterns. In *Journal of the Operational Research Society* (Vol. 57, Number 10, pp. 1256–1257). Nature Publishing Group. https://doi.org/10.1057/palgrave.jors.2602211

Purnamasari, D. I., Permadi, V. A., Saepudin, A., & Agusdin, R. P. (2023). Demand Forecasting for Improved Inventory Management in Small and Medium-Sized Businesses. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *12*(1), 56–66. https://doi.org/10.23887/janapati.v12i1.57144

Sana, J. K., Abedin, M. Z., Rahman, M. S., & Rahman, M. S. (2022). *Data transformation based optimized customer churn prediction model for the telecommunication industry*. https://doi.org/10.1371/journal.pone.0278095

Solis, A. O., Nicoletti, L., Mukhopadhyay, S., Agosteo, L., Delfino, A., & Sartiano, M. (n.d.). *INTERMITTENT DEMAND FORECASTING AND STOCK CONTROL: AN EMPIRICAL STUDY*.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495–503. https://doi.org/10.1057/palgrave.jors.2601841