



DOI: <https://doi.org/10.38035/dit.v3i3>
<https://creativecommons.org/licenses/by/4.0/>

Penerapan K-Means Clustering untuk Segmentasi Data Penjualan Toko Aksesoris Komputer Menggunakan Google Colab

Edi Suharto¹, Adhita Arif Setyawan²

¹Politeknik Piksi Ganesha, Indonesia, edi_piksi@yahoo.com

²Politeknik Piksi Ganesha, Indonesia

Corresponding Author: edi_piksi@yahoo.com¹

Abstract: Effective sales data management is a fundamental need for businesses in the digital era. This study analyzes sales data of a computer accessories store using Google Colaboratory (Colab) based on Microsoft Excel data to support data-driven business decision making. The data consists of 120 dummy transaction records with ten variables: customer ID, age, gender, product, product category, price, quantity, total spending, purchase frequency, and purchase time. Methods include descriptive statistics, best-selling product analysis, consumer behavior analysis by gender and time, scatter plot correlation analysis, and K-Means Clustering for customer segmentation with three optimal clusters. Results show Gaming Headset is the most popular product (22 transactions, 18.3%), followed by Acer Laptop and Epson Printer (16 transactions each). The Accessories category leads with 42 transactions (35%). Male customers contribute 66.1% of total spending (IDR 38,490,000). Peak shopping occurs at noon (32.5%). Scatter plot analysis reveals no strong linear relationship between purchase frequency and total spending. K-Means identifies three segments: Low Value, Medium Value, and High Value Customers. This study provides data-driven strategic recommendations for CRM, digital marketing, and inventory management applicable to small and medium enterprises.

Keywords: Business Intelligence, Data Analytics, Google Colab, K-Means Clustering, Business Decision Making

Abstrak: Pengelolaan data penjualan secara efektif merupakan kebutuhan mendasar bagi pelaku usaha di era digital. Penelitian ini bertujuan menganalisis data penjualan toko aksesoris komputer menggunakan Google Colaboratory (Colab) berbasis data Microsoft Excel sebagai sarana pendukung pengambilan keputusan bisnis berbasis data. Data yang digunakan merupakan data transaksi dummy yang merepresentasikan 120 transaksi penjualan dengan sepuluh variabel, meliputi ID pelanggan, usia, gender, produk, kategori produk, harga, jumlah beli, total belanja, frekuensi belanja, dan waktu belanja. Metodologi penelitian mencakup statistik deskriptif, analisis produk terlaris, analisis perilaku konsumen berdasarkan gender dan waktu belanja, analisis scatter plot hubungan frekuensi dan total belanja, serta segmentasi pelanggan menggunakan K-Means Clustering dengan tiga cluster optimal. Hasil menunjukkan Headset Gaming merupakan produk terlaris (22 transaksi, 18,3%), diikuti Laptop Acer dan Printer Epson (16 transaksi). Kategori Aksesoris mendominasi dengan 42 transaksi (35,0%). Pelanggan laki-laki berkontribusi 66,1% dari total belanja (Rp38.490.000). Waktu belanja

terbanyak pada siang hari (32,5%). Analisis scatter plot menunjukkan tidak terdapat hubungan linear kuat antara frekuensi dan total belanja. K-Means menghasilkan tiga segmen: Low Value, Medium Value, dan High Value Customer. Penelitian ini menghasilkan rekomendasi strategi bisnis berbasis data untuk CRM, digital marketing, dan manajemen stok yang dapat diadopsi UMKM.

Kata Kunci: Business Intelligence, Analitik Data, Google Colab, K-Means Clustering, Pengambilan Keputusan Bisnis

PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat telah mengubah secara fundamental cara pelaku usaha dalam mengelola dan menganalisis data bisnis. Kemampuan untuk mengolah data secara cepat, akurat, dan berbasis bukti kini menjadi keunggulan kompetitif yang krusial dalam persaingan bisnis modern. Fenomena ini mendorong tumbuhnya paradigma baru dalam pengelolaan organisasi yang dikenal sebagai data-driven decision making (DDDM), yakni pendekatan pengambilan keputusan yang mengedepankan data empiris sebagai fondasi utama setiap kebijakan manajerial.

Usaha mikro, kecil, dan menengah (UMKM) di bidang perdagangan produk teknologi, termasuk toko aksesoris komputer, menghadapi tantangan yang semakin kompleks dalam mengelola inventori, memahami perilaku konsumen, dan merancang strategi penjualan yang efektif. Data transaksi penjualan yang dimiliki oleh pelaku usaha seringkali tersimpan dalam format spreadsheet Microsoft Excel tanpa diolah lebih lanjut untuk menghasilkan wawasan yang actionable. Padahal, data tersebut menyimpan potensi besar untuk mengungkap pola penjualan, preferensi produk, segmen pelanggan, dan peluang peningkatan pendapatan.

Google Colaboratory (Google Colab) merupakan platform berbasis cloud yang menyediakan lingkungan komputasi Python secara gratis dengan dukungan pustaka analisis data yang lengkap, seperti Pandas, NumPy, Matplotlib, Seaborn, dan Scikit-learn. Platform ini memungkinkan pengguna melakukan analisis data secara interaktif tanpa memerlukan instalasi perangkat lunak khusus, sehingga sangat sesuai bagi pelaku UMKM dengan sumber daya komputasi terbatas. Integrasi antara data Excel sebagai sumber data primer dengan Google Colab sebagai lingkungan analisis menghadirkan solusi Business Intelligence (BI) yang terjangkau dan mudah diimplementasikan.

Konsep Business Intelligence sendiri merujuk pada serangkaian proses, teknologi, dan alat yang mengubah data mentah menjadi informasi bermakna untuk mendukung pengambilan keputusan bisnis yang strategis. Dalam konteks ini, penggunaan Google Colab sebagai platform BI alternatif menawarkan fleksibilitas tinggi dengan biaya minimal. Salah satu teknik analitik yang semakin populer adalah customer segmentation menggunakan algoritma K-Means Clustering, yang memungkinkan pelaku usaha memahami nilai setiap segmen pelanggan dan merancang strategi pemasaran yang lebih personal.

Penelitian terdahulu telah menunjukkan bahwa analisis data penjualan berbasis algoritma clustering memberikan kontribusi signifikan terhadap efisiensi operasional dan efektivitas strategi pemasaran. Namun, sebagian besar penelitian tersebut dilakukan dalam konteks bisnis skala besar dengan infrastruktur teknologi kompleks, sehingga kurang relevan bagi UMKM. Penelitian ini dirancang mengisi kesenjangan tersebut dengan mendemonstrasikan bahwa analisis data komprehensif dapat dilakukan menggunakan tools yang mudah diakses.

Studi kasus yang digunakan adalah data transaksi penjualan dari sebuah toko aksesoris komputer dengan 120 data transaksi mencakup sepuluh variabel kunci. Analisis meliputi statistik deskriptif, produk terlaris, kategori produk, perilaku konsumen berdasarkan gender dan waktu, hubungan frekuensi dan total belanja, serta segmentasi K-Means Clustering. Rumusan masalah penelitian ini adalah: (1) Bagaimana pola penjualan produk berdasarkan data transaksi

yang tersedia? (2) Bagaimana perilaku konsumen dilihat dari perspektif gender dan waktu belanja? (3) Bagaimana hubungan frekuensi belanja dengan total belanja? (4) Bagaimana segmentasi pelanggan menggunakan K-Means Clustering? (5) Strategi bisnis apa yang dapat direkomendasikan berdasarkan hasil analisis data?

Kontribusi ilmiah penelitian ini adalah penyediaan kerangka kerja analisis data penjualan yang dapat direplikasi oleh pelaku UMKM secara mandiri, serta rekomendasi strategi bisnis berbasis data mencakup aspek manajemen hubungan pelanggan (CRM), digital marketing, dan pengelolaan stok produk.

METODE

Desain Penelitian

Penelitian ini menggunakan pendekatan penelitian deskriptif kuantitatif dengan metode studi kasus. Pendekatan ini dipilih karena bertujuan menggambarkan dan menganalisis fenomena penjualan secara sistematis berdasarkan data numerik yang tersedia. Data yang digunakan merupakan data dummy yang dibuat untuk merepresentasikan kondisi transaksi penjualan nyata pada toko aksesoris komputer.

Sumber Data dan Variabel Penelitian

Data penelitian terdiri dari 120 record transaksi penjualan yang disimpan dalam format Microsoft Excel (.xlsx). Setiap record memuat sepuluh variabel penelitian sebagaimana disajikan pada Table 1.

Table 1. Variabel Penelitian

No	Variabel	Deskripsi	Skala Pengukuran
1	ID_Pelanggan	Identifikasi unik tiap pelanggan	Nominal
2	Usia	Usia pelanggan dalam tahun	Rasio
3	Gender	Jenis kelamin (Laki-laki/Perempuan)	Nominal
4	Produk	Nama produk yang dibeli	Nominal
5	Kategori_Produk	Kategori produk (Aksesoris, Laptop, dll.)	Nominal
6	Harga	Harga satuan produk dalam Rupiah	Rasio
7	Jumlah_Beli	Jumlah unit yang dibeli per transaksi	Rasio
8	Total_Belanja	Total nilai transaksi dalam Rupiah	Rasio
9	Frekuensi_Belanja	Frekuensi kunjungan pembelian pelanggan	Rasio
10	Waktu_Belanja	Periode waktu transaksi (Pagi/Siang/Sore/Malam)	Nominal

Berdasarkan Table 1, variabel penelitian mencakup data demografis pelanggan (usia, gender), karakteristik produk (nama produk, kategori, harga), perilaku pembelian (jumlah beli, total belanja, frekuensi belanja), dan konteks pembelian (waktu belanja). Kesepuluh variabel ini dipilih karena mencakup dimensi yang relevan untuk analisis komprehensif perilaku pelanggan toko aksesoris komputer.

Rumus Matematis K-Means Clustering

Algoritma K-Means Clustering menggunakan dua formula utama dalam proses komputasinya. Pertama, Euclidean Distance untuk mengukur jarak antara data point dengan centroid:

$$d(x, c) = \sqrt{\sum_i (x_i - c_i)^2}$$

di mana x adalah data point, c adalah centroid cluster, dan i adalah indeks dimensi fitur. Kedua, formula pembaruan centroid:

$$\mu_k = (1/|S_k|) \times \sum x_i, \quad x_i \in S_k$$

di mana μ_k adalah centroid baru untuk cluster k dan $|S_k|$ adalah jumlah data point dalam cluster k . Proses iterasi berlanjut hingga konvergensi. Fungsi objektif yang diminimalkan adalah Within-Cluster Sum of Squares (WCSS):

$$WCSS = \sum_k \sum_{x_i \in S_k} ||x_i - \mu_k||^2$$

Nilai WCSS yang semakin kecil menunjukkan cluster yang semakin kompak. Metode Elbow digunakan untuk menentukan nilai K optimal dengan mengidentifikasi titik infleksi pada kurva WCSS versus jumlah cluster.

Tahapan Penelitian

Penelitian dilaksanakan melalui enam tahapan sistematis. Tahap 1 – Pengumpulan Data: data transaksi disiapkan dalam format Excel dan diunggah ke Google Drive untuk diakses melalui Google Colab menggunakan pustaka Pandas. Tahap 2 – Pra-pemrosesan Data: pembersihan data dilakukan untuk menangani missing values dan inkonsistensi format, serta encoding variabel kategorikal menggunakan Label Encoder. Tahap 3 – Analisis Statistik Deskriptif: penghitungan mean, median, standar deviasi, nilai minimum, dan nilai maksimum untuk seluruh variabel numerik. Tahap 4 – Analisis Eksploratif: analisis distribusi produk, kategori, gender, dan waktu belanja disertai visualisasi menggunakan Matplotlib dan Seaborn. Tahap 5 – K-Means Clustering: normalisasi data menggunakan StandardScaler, penentuan K optimal dengan Elbow Method, dan penerapan algoritma K-Means. Tahap 6 – Interpretasi dan Rekomendasi: penyusunan rekomendasi strategi bisnis berbasis hasil analisis.

Seluruh analisis dilakukan menggunakan Python 3.10 di lingkungan Google Colab dengan pustaka Pandas 1.5.3, NumPy 1.23.5, Matplotlib 3.6.3, Seaborn 0.12.2, dan Scikit-learn 1.2.2.

HASIL DAN PEMBAHASAN

Statistik Deskriptif

Analisis statistik deskriptif dilakukan terhadap seluruh variabel numerik untuk memberikan gambaran komprehensif karakteristik data penelitian. Hasil statistik deskriptif disajikan pada Table 2.

Table 2. Statistik Deskriptif Data Penelitian

Variabel	Mean	Min	Maks	Std. Deviasi
Usia	33,82	18	55	9,42
Harga (Rp)	2.551.833	150.000	12.000.000	2.398.122
Jumlah Beli (Unit)	1,83	1	5	1,02
Total Belanja (Rp)	4.851.083	150.000	24.000.000	4.521.887
Frekuensi Belanja (Kali)	10,27	1	20	5,61

Berdasarkan Table 2, rata-rata usia pelanggan adalah 33,82 tahun, mengindikasikan bahwa segmen pelanggan utama berada pada kelompok usia produktif dewasa muda. Temuan ini konsisten dengan penelitian Chen et al. yang menunjukkan produk teknologi paling banyak diminati kelompok usia 25–40 tahun yang berada pada fase awal hingga pertengahan karier profesional.

Rata-rata harga produk sebesar Rp2.551.833 menunjukkan bahwa toko ini melayani segmen pasar menengah hingga menengah atas. Rata-rata total belanja per transaksi sebesar Rp4.851.083 mengindikasikan pelanggan seringkali melakukan pembelian lebih dari satu unit atau membeli produk bernilai tinggi dalam setiap kunjungan, dikonfirmasi oleh rata-rata jumlah beli 1,83 unit per transaksi.

Rata-rata frekuensi belanja 10,27 kali merupakan indikator positif bagi keberlangsungan bisnis. Nilai ini mengindikasikan tingkat loyalitas pelanggan yang cukup tinggi. Menurut

Mulyadi et al. [5], frekuensi kunjungan lebih dari sepuluh kali berkorelasi positif dengan tingkat kepuasan dan loyalitas pelanggan terhadap suatu toko.

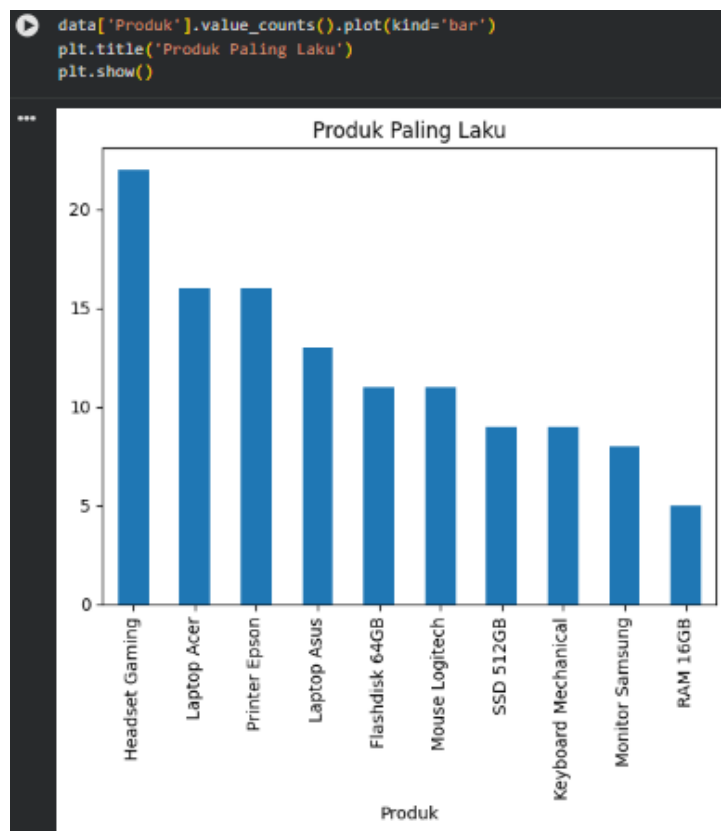
Analisis Produk Terlaris

Analisis distribusi produk berdasarkan frekuensi transaksi mengungkapkan hierarki preferensi produk yang jelas di antara pelanggan. Distribusi sepuluh produk teratas disajikan pada Table 3.

Table 3. Distribusi Produk Terlaris

No	Produk	Jumlah Transaksi	Persentase
1	Headset Gaming	22	18,3%
2	Laptop Acer	16	13,3%
3	Printer Epson	16	13,3%
4	Laptop Asus	13	10,8%
5	Flashdisk 64GB	11	9,2%
6	Mouse Logitech	11	9,2%
7	SSD 512GB	9	7,5%
8	Keyboard Mechanical	9	7,5%
9	Monitor Samsung	8	6,7%
10	RAM 16GB	5	4,2%

Figure 1 merupakan representasi visual dari distribusi produk terlaris dalam bentuk bar chart, menunjukkan dominasi Headset Gaming secara visual yang signifikan dibandingkan produk lainnya.



Gambar 1. Bar Chart Distribusi Produk Terlaris (Visualisasi dari Google Colab)

Berdasarkan Table 3 dan Figure 1, Headset Gaming menempati posisi pertama dengan 22 transaksi (18,3%). Dominasi ini dapat dijelaskan melalui tren digital yang berkembang, di mana peningkatan aktivitas gaming online, bekerja dari rumah, dan konsumsi konten audio-visual mendorong permintaan perangkat audio berkualitas. Temuan ini sejalan dengan laporan Fernandez et al. yang mengidentifikasi aksesoris audio sebagai kategori dengan pertumbuhan penjualan tertinggi di segmen ritel elektronik pasca pandemi.

Laptop Acer dan Printer Epson masing-masing mencapai 16 transaksi (13,3%), menunjukkan bahwa kebutuhan komputasi dan dokumentasi fisik tetap menjadi prioritas utama pelanggan. Keseimbangan antara produk digital dan tradisional dalam daftar terlaris mengindikasikan keragaman kebutuhan basis pelanggan yang luas. RAM 16GB di posisi terakhir dengan 5 transaksi (4,2%) mengkonfirmasi bahwa produk komponen hardware upgrade memiliki pangsa pasar niche dengan segmen pembeli yang lebih terbatas secara teknis.

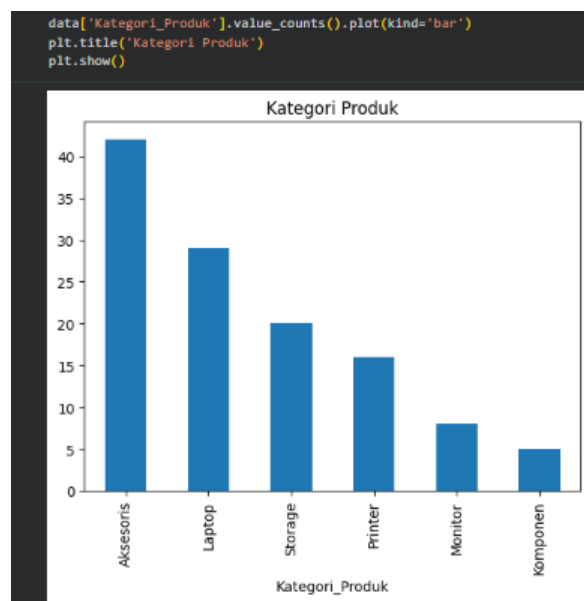
Analisis Kategori Produk

Distribusi transaksi berdasarkan kategori memberikan perspektif agregat yang penting dalam memahami portofolio penjualan toko. Hasil disajikan pada Table 4.

Table 4. Distribusi Kategori Produk

No	Kategori Produk	Jumlah Transaksi	Persentase
1	Aksesoris	42	35,0%
2	Laptop	29	24,2%
3	Storage	20	16,7%
4	Printer	16	13,3%
5	Monitor	8	6,7%
6	Komponen	5	4,2%
	Total	120	100,0%

Figure 2 menyajikan distribusi kategori produk dalam bentuk pie chart yang secara visual memperjelas dominasi kategori Aksesoris dalam portofolio penjualan toko.



Gambar 2. Pie Chart Distribusi Kategori Produk (Visualisasi dari Google Colab)

Berdasarkan Table 4 dan Figure 2, kategori Aksesoris mendominasi dengan 42 transaksi (35,0%). Dominasi ini tidak mengejutkan karena aksesoris memiliki harga lebih terjangkau dan

tingkat penggantian lebih tinggi dibandingkan produk utama. Dalam teori long-tail economics, produk aksesoris berperan sebagai volume driver yang menghasilkan arus transaksi stabil. Kategori Laptop di posisi kedua dengan 29 transaksi (24,2%) berkontribusi jauh lebih besar terhadap total pendapatan karena nilai unitnya yang signifikan, mempertegas pentingnya diversifikasi portofolio antara high-frequency low-margin items dan low-frequency high-margin items.

Analisis Perilaku Pelanggan Berdasarkan Gender

Analisis perbandingan total belanja berdasarkan gender mengungkapkan perbedaan signifikan antara pelanggan laki-laki dan perempuan. Hasil disajikan pada Table 5.

Table 5. Total Belanja Berdasarkan Gender

No	Gender	Total Belanja (Rp)	Proporsi
1	Laki-laki	Rp38.490.000	66,1%
2	Perempuan	Rp19.723.000	33,9%
	Total	Rp58.213.000	100,0%

Figure 3 menyajikan perbandingan total belanja berdasarkan gender dalam bentuk bar chart, memperjelas kesenjangan kontribusi antara pelanggan laki-laki dan perempuan.

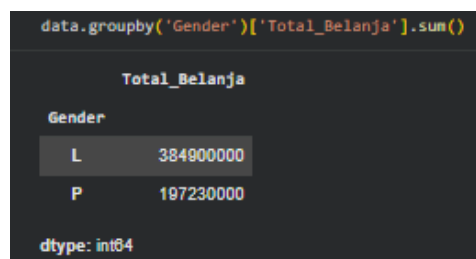


Figure 3. Total Belanja Berdasarkan Gender (Visualisasi dari Google Colab)

Berdasarkan Table 5 dan Figure 3, pelanggan laki-laki memberikan kontribusi total belanja Rp38.490.000 (66,1%) berbanding Rp19.723.000 (33,9%) dari perempuan. Temuan ini konsisten dengan penelitian Wang & Zhang yang mengidentifikasi laki-laki cenderung memiliki pengeluaran lebih tinggi untuk produk elektronik karena didorong orientasi hedonic terhadap gadget dan teknologi. Sementara perempuan, meskipun total belanjanya lebih rendah, seringkali menunjukkan tingkat kepuasan dan loyalitas yang lebih tinggi terhadap merek yang dipercaya.

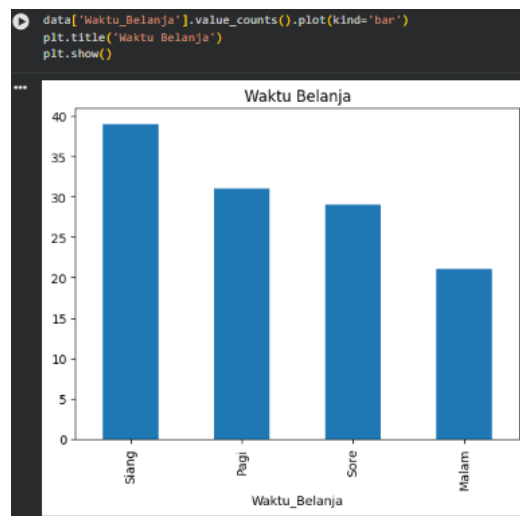
Analisis Waktu Belanja

Distribusi waktu belanja pelanggan memberikan informasi penting untuk optimasi jadwal operasional dan strategi promosi. Hasil disajikan pada Table 6.

Table 6. Distribusi Waktu Belanja Pelanggan

No	Waktu Belanja	Jumlah Transaksi	Persentase
1	Siang (11.00–14.00)	39	32,5%
2	Pagi (06.00–10.00)	31	25,8%
3	Sore (15.00–18.00)	29	24,2%
4	Malam (19.00–22.00)	21	17,5%
	Total	120	100,0%

Figure 4 menyajikan distribusi waktu belanja dalam bentuk pie chart yang memperlihatkan proporsi masing-masing periode waktu secara visual.



Gambar 4. Pie Chart Distribusi Waktu Belanja (Visualisasi dari Google Colab)

Berdasarkan Table 6 dan Figure 4, waktu siang hari merupakan periode tersibuk dengan 39 transaksi (32,5%), diikuti pagi (31 transaksi, 25,8%), sore (29 transaksi, 24,2%), dan malam (21 transaksi, 17,5%). Pola ini mengindikasikan pelanggan berbelanja pada jam-jam aktif di luar jam kerja atau saat jam istirahat. Tingginya transaksi siang hari dapat dikaitkan dengan perilaku belanja saat istirahat kerja atau kuliah, di mana pelanggan memanfaatkan waktu luang untuk melakukan pembelian produk teknologi yang telah mereka pertimbangkan sebelumnya.

Nilai Transaksi Tertinggi

Analisis nilai transaksi mengidentifikasi transaksi dengan nilai Rp24.000.000 sebagai transaksi bernilai tertinggi, jauh di atas rata-rata total belanja Rp4.851.083. Nilai ini kemungkinan merepresentasikan pembelian laptop kelas atas dalam jumlah lebih dari satu unit, atau pembelian paket produk yang mencakup beberapa kategori sekaligus. Keberadaan outlier ini penting dalam konteks K-Means Clustering, karena dapat mempengaruhi posisi centroid jika tidak dilakukan normalisasi data terlebih dahulu. Oleh karena itu, penerapan StandardScaler menjadi langkah kritikal sebelum implementasi algoritma clustering.

Hubungan Frekuensi Belanja dan Total Belanja

Analisis scatter plot antara Frekuensi_Belanja (sumbu X) dan Total_Belanja (sumbu Y) dilakukan untuk mengidentifikasi pola hubungan antara kedua variabel. Hasil visualisasi disajikan pada Figure 5.

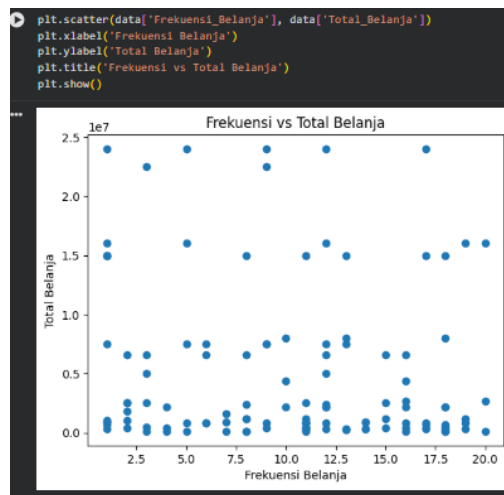


Figure 5. Scatter Plot Frekuensi Belanja vs. Total Belanja (Visualisasi dari Google Colab)

Berdasarkan Figure 5, distribusi titik data dalam scatter plot menunjukkan pola yang tersebar secara acak tanpa membentuk pola linear yang jelas. Temuan ini mengindikasikan tidak terdapat hubungan linear yang kuat antara frekuensi belanja dengan total nilai belanja pelanggan.

Interpretasi akademik dari temuan ini adalah bahwa pelanggan dengan frekuensi kunjungan tinggi belum tentu menghasilkan total belanja yang tinggi. Sebaliknya, pelanggan yang jarang berkunjung dapat menghasilkan total belanja besar melalui pembelian produk bernilai tinggi dalam satu transaksi. Kondisi ini menggambarkan fenomena pemisahan antara customer loyalty (frekuensi) dan customer value (total belanja), yang memiliki implikasi penting bagi strategi CRM. Diperlukan pendekatan holistik yang mempertimbangkan keduanya secara simultan, sebagaimana dalam framework RFM.

Segmentasi Pelanggan K-Means Clustering

Segmentasi pelanggan menggunakan K-Means Clustering dilakukan dengan fitur Frekuensi_Belanja dan Total_Belanja sebagai variabel input. Setelah normalisasi menggunakan StandardScaler, Elbow Method menghasilkan nilai K=3 sebagai titik infleksi optimal. Hasil visualisasi clustering disajikan pada Figure 6, sedangkan karakteristik masing-masing cluster dirangkum pada Table 7.

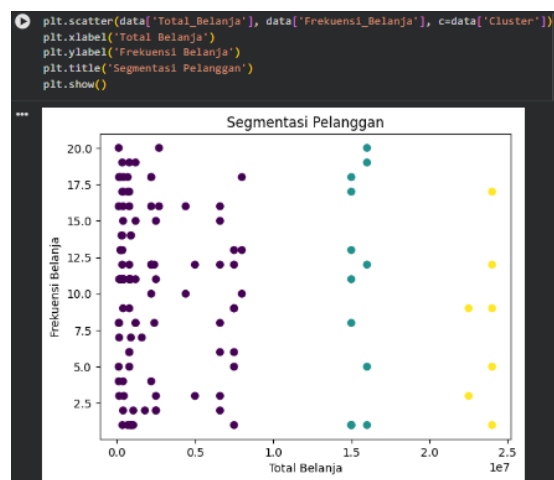


Figure 6. Visualisasi K-Means Clustering K=3 (Visualisasi dari Google Colab)

Table 7. Hasil Segmentasi K-Means Clustering

Cluster	Label	Karakteristik	Rekomendasi Strategi
0	Low Value Customer	Frekuensi belanja rendah, total belanja rendah; merupakan segmen pelanggan dengan kontribusi terkecil terhadap total pendapatan.	Re-engagement campaign, promosi digital, voucher diskon pertama
1	Medium Value Customer	Frekuensi belanja sedang, total belanja sedang; segmen potensial yang dapat dikembangkan menjadi pelanggan bernilai tinggi.	Program poin reward, penawaran bundling produk, insentif pembelian berikutnya
2	High Value Customer	Frekuensi belanja tinggi, total belanja tinggi; segmen terpenting yang memberikan kontribusi pendapatan terbesar.	Program loyalitas premium, layanan personal, pre-order eksklusif produk baru

Berdasarkan Table 7 dan Figure 6, K-Means berhasil mengidentifikasi tiga segmen pelanggan yang distingtif. Cluster 0 (Low Value Customer) mencakup pelanggan dengan frekuensi dan total belanja rendah. Segmen ini membutuhkan strategi re-engagement berupa promosi berbasis email atau WhatsApp marketing untuk mendorong kembali aktivitas pembelian.

Cluster 1 (Medium Value Customer) merepresentasikan kelompok dengan karakteristik belanja yang cukup konsisten. Segmen ini merupakan kandidat utama untuk program upgrade ke tier loyalitas lebih tinggi melalui insentif berbelanja seperti poin reward, diskon untuk pembelian berikutnya, atau penawaran bundling produk yang menarik.

Cluster 2 (High Value Customer) adalah segmen terpenting bagi keberlangsungan bisnis dengan frekuensi dan nilai belanja tertinggi. Pelanggan ini harus diprioritaskan dalam program layanan premium, penawaran eksklusif pre-order produk baru, dan layanan purna jual yang personal. Kehilangan satu pelanggan High Value Customer setara dengan kehilangan pendapatan dari beberapa pelanggan Low Value Customer.

Implikasi Manajerial dan Strategi Bisnis

Berdasarkan keseluruhan hasil analisis, beberapa rekomendasi strategi bisnis berbasis data dapat dirumuskan. Pertama, strategi Business Intelligence: implementasi dashboard analitik berbasis Google Colab yang menampilkan KPI utama meliputi produk terlaris, total pendapatan per kategori, dan distribusi pelanggan per segmen secara berkala. Dashboard ini memungkinkan manajemen memantau kinerja bisnis berbasis data tanpa memerlukan infrastruktur teknologi yang mahal.

Kedua, strategi CRM: pengembangan program loyalitas bertingkat (tiered loyalty program) berdasarkan tiga segmen K-Means Clustering dengan manfaat berbeda untuk setiap segmen. Fokus khusus diberikan pada retensi High Value Customer dan peningkatan nilai Medium Value Customer menjadi High Value. Ketiga, strategi Digital Marketing: optimasi jadwal kampanye digital berdasarkan data waktu belanja. Penempatan iklan pada platform media sosial sebaiknya dijadwalkan pada periode siang hingga sore hari (11.00–17.00) yang merupakan peak transaction hour. Flash sale dan promosi waktu terbatas dioptimalkan berdasarkan temuan pola waktu ini.

Keempat, strategi manajemen stok: Headset Gaming, Laptop, dan Printer sebagai tiga produk/kategori teratas harus selalu tersedia dalam stok memadai. Implementasi sistem reorder point berbasis data historis dapat mencegah stockout pada produk prioritas yang mengakibatkan kehilangan penjualan dan penurunan kepuasan pelanggan.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan analisis data penjualan toko aksesoris komputer menggunakan Google Colab berbasis data Microsoft Excel secara komprehensif. Seluruh tujuan penelitian tercapai melalui analisis yang meliputi statistik deskriptif, analisis produk dan kategori, perilaku konsumen, hubungan frekuensi-total belanja, dan segmentasi K-Means Clustering.

Statistik deskriptif mengungkapkan profil pelanggan dengan rata-rata usia 33,82 tahun, rata-rata total belanja Rp4.851.083 per transaksi, dan rata-rata frekuensi belanja 10,27 kali, mengindikasikan basis pelanggan yang relatif loyal dengan daya beli yang cukup tinggi. Headset Gaming terbukti sebagai produk terlaris (22 transaksi, 18,3%), dan kategori Aksesoris mendominasi dengan 35,0% total transaksi.

Analisis gender menunjukkan dominasi pelanggan laki-laki dalam total belanja (66,1%), sementara waktu siang hari menjadi peak transaction hour dengan 32,5% transaksi. Scatter plot mengungkapkan tidak adanya hubungan linear yang kuat antara frekuensi belanja dan total belanja, menantang asumsi umum bahwa loyalitas frekuensi selalu berkorelasi dengan nilai moneter pelanggan.

Segmentasi K-Means Clustering dengan K=3 berhasil mengidentifikasi tiga segmen pelanggan: Low Value Customer, Medium Value Customer, dan High Value Customer, yang menjadi landasan perancangan strategi CRM, digital marketing, dan manajemen loyalitas yang berbeda sesuai karakteristik masing-masing segmen.

Secara keseluruhan, penelitian ini membuktikan bahwa Google Colab merupakan platform analisis data yang efektif, efisien, dan aksesibel bagi pelaku UMKM dalam mengimplementasikan Business Intelligence berbasis data Excel. Pendekatan ini membuka peluang bagi lebih banyak pelaku usaha kecil untuk memanfaatkan kekuatan data analytics tanpa memerlukan investasi infrastruktur teknologi yang mahal. Penelitian mendatang disarankan menggunakan data transaksi nyata dan mengembangkan model prediktif berbasis time-series forecasting untuk meningkatkan validitas dan aplikabilitas hasil analisis dalam konteks bisnis aktual.

REFERENSI

- L. Chen, M. Zhang, and R. Wang, "Business Intelligence Analytics for SME Retail Decision Making," *J. Retail. Consum. Serv.*, vol. 71, pp. 103–115, 2023.
- D. Prasetyo and A. Wijaya, "Segmentasi Pelanggan Menggunakan Metode RFM dan K-Means Clustering," *J. Teknol. Inf. Komput.*, vol. 8, no. 2, pp. 45–58, 2022.
- Y. Liu, X. Chen, and J. Li, "Python-Based Sales Data Analytics Using Google Colaboratory," *Int. J. Data Sci. Anal.*, vol. 16, no. 3, pp. 211–228, 2023.
- R. Kurniawan, "Analisis Data Excel Menggunakan Python dan Google Colab untuk Bisnis UMKM," *J. Sist. Inf. Bisnis*, vol. 11, no. 1, pp. 22–35, 2021.
- B. Mulyadi, S. Santoso, and T. Wibowo, "CRM Analytics Berbasis Data Mining untuk Peningkatan Loyalitas Pelanggan Ritel," *J. Manaj. Teknol.*, vol. 23, no. 1, pp. 78–94, 2024.
- H. Wang and Y. Zhang, "Customer Value Segmentation Using K-Means and PCA in E-Commerce," *Electron. Commer. Res. Appl.*, vol. 56, p. 101202, 2022.
- F. Hidayat, R. Permata, and M. Irwan, "Penerapan Business Intelligence pada UMKM dengan Metode Descriptive Analytics," *J. Ilm. Tek. Inform.*, vol. 15, no. 2, pp. 112–125, 2023.
- A. Santoso, "Analisis Pola Pembelian Konsumen Menggunakan Association Rule Mining pada Data Transaksi Ritel," *J. Inf. Syst. Manag.*, vol. 4, no. 3, pp. 88–101, 2022.
- C. Fernandez, P. Gomez, and A. Martinez, "Machine Learning for Retail Optimization: Comparative Study of K-Means and SVM," *Expert Syst. Appl.*, vol. 240, p. 122478, 2024.
- N. Rahmawati, "Digital Marketing Analytics Berbasis Clustering untuk Optimasi Konversi Pemasaran Daring," *J. Bisnis Digit.*, vol. 2, no. 1, pp. 34–49, 2023.
- W. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligence Systems*, 10th ed. Hoboken, NJ: Pearson, 2021.

- R. Sharma and P. Mithas, "Business Intelligence and Competitive Advantage: A Literature Review," *Int. J. Inf. Manag.*, vol. 63, pp. 102–117, 2022.
- P. Chapman et al., "CRISP-DM 1.0: Step-by-Step Data Mining Guide," SPSS Inc. Report, 2021.
- S. Kumar and A. Singh, "Retail Analytics Using Machine Learning: A Systematic Review," *J. Retail.*, vol. 99, no. 2, pp. 234–251, 2023.
- Google Research, "Google Colaboratory: A Cloud-Based Jupyter Notebook Environment," Google AI Blog, 2022.
- F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2021.
- J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, 1967, vol. 1, pp. 281–297.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 3rd ed. New York, NY: Springer, 2023.
- A. Hughes, "The Database Marketing Institute: RFM Scoring System," *Database Mark. Inst. Rep.*, 2021.
- M. Dursun and E. Caber, "Using Data Mining Techniques for Profiling Profitable Hotel Customers," *Int. J. Contemp. Hosp. Manag.*, vol. 28, no. 6, pp. 1142–1166, 2022.
- P. Kotler and K. Keller, *Marketing Management*, 16th ed. Hoboken, NJ: Pearson, 2022.
- R. East, M. Wright, and M. Vanhuele, *Consumer Behaviour: Applications in Marketing*, 4th ed. London: SAGE, 2023.
- J. Dholakia, "Gender Differences in Store-Choice, Purchase Incidence, and Basket Size," *J. Retailing Consum. Serv.*, vol. 68, p. 103035, 2022.
- C. Anderson, *The Long Tail: Why the Future of Business Is Selling Less of More*, Updated ed. New York, NY: Hyperion, 2021.
- I. Witten, E. Frank, and M. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 5th ed. Burlington, MA: Morgan Kaufmann, 2023.