**Exploratory Data Analysis pada Pola Produk dan Penjual di Platform E-commerce**

**M Islam Al Wali**1**\***

1,2,3 Institut Teknologi Nasional Malang, Indonesia

1[islam.alwali42@gmail.com](mailto:adenaryawardana@gmail.com)

|  |  |
| --- | --- |
| **Histori Artikel:**  Diajukan: Disetujui:  Dipublikasi: | **Abstrak**  Penelitian ini bertujuan untuk melakukan Exploratory Data Analysis (EDA) terhadap dataset produk e-commerce guna memahami pola distribusi kategori dan aktivitas penjual di platform daring. Analisis dilakukan menggunakan dataset data\_products\_id\_small (2).csv yang berisi informasi tentang produk, toko, dan kategori. Dengan pendekatan EDA, penelitian ini berfokus pada identifikasi kategori produk yang paling dominan, sub-kategori terpopuler, serta toko dengan jumlah produk terbanyak. Pendekatan ini penting untuk memahami karakteristik dasar dari data yang tersedia sebelum dilakukan analisis lanjutan seperti prediksi penjualan atau rekomendasi produk.  Tahapan analisis meliputi pembersihan data, eksplorasi atribut kategorikal, visualisasi distribusi, serta analisis korelasi antar atribut numerik. Data dianalisis menggunakan bahasa pemrograman Python dengan pustaka pandas, seaborn, dan matplotlib untuk membantu proses visualisasi. Dari hasil analisis ditemukan bahwa sebagian besar produk berada dalam kategori Elektronik, khususnya sub-kategori Konsol Game, dengan beberapa toko menampilkan jumlah listing produk yang jauh lebih besar dibandingkan toko lainnya. Hal ini menunjukkan adanya konsentrasi produk dalam kategori tertentu dan aktivitas tinggi dari penjual tertentu di dalam platform.  Penelitian ini memberikan gambaran umum mengenai pola penyebaran produk dan aktivitas penjual di e-commerce, yang dapat menjadi dasar bagi analisis perilaku konsumen, strategi pemasaran, dan pengelolaan inventori. Melalui EDA, pemahaman terhadap struktur data dan hubungan antar variabel dapat membantu dalam pengambilan keputusan yang lebih efektif. Hasil analisis ini juga dapat digunakan sebagai referensi dalam pengembangan sistem rekomendasi atau pengelompokan penjual di masa mendatang. |
| **Kata Kunci:** E-commerce, Produk, Penjual, Kategori, Exploratory Data Analysis, Marketplace |
| ***Digital Transformation Technology (Digitech)*** *is an This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).* |

# **PENDAHULUAN**

Perkembangan teknologi digital telah mendorong transformasi besar dalam dunia perdagangan, terutama melalui munculnya platform e-commerce. Fenomena ini mengubah cara masyarakat membeli dan menjual produk, dari sistem konvensional menjadi transaksi daring yang serba cepat dan efisien. Berdasarkan laporan We Are Social (2024), lebih dari 2,6 miliar orang di dunia telah melakukan pembelian melalui platform e-commerce, dan angka tersebut terus meningkat setiap tahun. Kondisi ini menunjukkan bahwa sektor perdagangan digital memainkan peran penting dalam perekonomian global serta membuka peluang baru bagi pelaku usaha untuk memperluas jangkauan pasar mereka.

Namun, pesatnya pertumbuhan ini juga menimbulkan tantangan baru, terutama dalam hal memahami perilaku dan pola penjualan di platform e-commerce. Banyaknya produk dan toko yang beroperasi secara bersamaan menghasilkan volume data yang sangat besar dan kompleks. Oleh karena itu, diperlukan pendekatan analisis data yang efektif untuk memahami karakteristik produk, distribusi kategori, serta aktivitas penjual. Salah satu metode yang dapat digunakan adalah Exploratory Data Analysis (EDA), yang memungkinkan peneliti untuk menggali informasi dan menemukan pola tersembunyi dari data tanpa memerlukan model prediktif yang rumit.

Penelitian ini bertujuan untuk melakukan analisis eksploratif terhadap dataset produk e-commerce yang berisi informasi mengenai produk, toko, serta kategori utama dan subkategori. Analisis dilakukan untuk mengidentifikasi kategori produk yang paling dominan, toko dengan jumlah produk terbanyak, dan distribusi keseluruhan produk di platform. Dengan penerapan EDA, diharapkan penelitian ini dapat memberikan wawasan awal mengenai struktur data e-commerce dan menjadi dasar bagi penelitian lanjutan, seperti analisis tren penjualan, perilaku konsumen, atau sistem rekomendasi produk di masa depan.

# **STUDI LITERATUR**

Konsep Exploratory Data Analysis (EDA) pertama kali diperkenalkan oleh John Tukey pada tahun 1977 sebagai pendekatan eksploratif untuk memahami data secara mendalam sebelum dilakukan pemodelan statistik yang kompleks. Menurut Tukey (1977), EDA berfokus pada penggunaan visualisasi, statistik deskriptif, dan identifikasi pola untuk menemukan struktur tersembunyi dalam data. Tahapan ini penting untuk memastikan kualitas data, mendeteksi anomali, serta memperoleh pemahaman awal terhadap karakteristik dataset. Han, Kamber, & Pei (2011) juga menegaskan bahwa EDA merupakan bagian penting dalam proses data preprocessing, karena membantu menyiapkan data agar layak digunakan dalam tahap analisis lanjutan seperti machine learning atau predictive modeling.

Dalam konteks e-commerce, EDA memiliki peran strategis dalam menganalisis produk, kategori, dan perilaku penjual. Chen et al. (2020) menjelaskan bahwa eksplorasi data penjualan online dapat membantu pelaku usaha memahami tren pasar dan preferensi konsumen secara real time. Dengan memanfaatkan analisis deskriptif, perusahaan dapat mengidentifikasi kategori produk yang paling laris, sub-kategori dengan pertumbuhan cepat, serta toko yang paling aktif dalam menambah listing produk. Hal ini sejalan dengan penelitian Xu & Li (2019) yang menunjukkan bahwa analisis eksploratif pada data toko online dapat meningkatkan efisiensi pengelolaan stok dan strategi pemasaran digital.

Menurut Gupta & Kohli (2019), e-commerce analytics merupakan penerapan prinsip-prinsip analisis data untuk mendukung keputusan bisnis berbasis data (data-driven decision making). Melalui visualisasi seperti bar chart, pie chart, dan heatmap correlation, peneliti dapat menggambarkan distribusi produk, perbandingan antar kategori, serta hubungan antar variabel numerik. Visualisasi ini tidak hanya membantu dalam menemukan pola tersembunyi, tetapi juga menjadi dasar pengembangan model prediksi penjualan atau rekomendasi produk di tahap selanjutnya.

Selain itu, riset oleh Lim & Tan (2021) menunjukkan bahwa pemahaman terhadap pola distribusi produk dan aktivitas penjual merupakan langkah awal yang krusial dalam transformasi digital bisnis retail. Dengan menggunakan EDA, perusahaan dapat menilai performa toko secara visual dan menyusun strategi peningkatan visibilitas produk. Oleh karena itu, penerapan Exploratory Data Analysis pada dataset produk e-commerce memberikan nilai tambah dalam memahami dinamika pasar dan perilaku pelaku bisnis secara menyeluruh.

**METODE**

Penelitian ini menggunakan metode kuantitatif eksploratif dengan pendekatan Exploratory Data Analysis (EDA). Analisis dilakukan tanpa menggunakan model prediktif, melainkan melalui eksplorasi visual dan statistik deskriptif untuk menemukan pola hubungan antar variabel dalam dataset. Tujuan utama metode ini adalah untuk memahami struktur data, mendeteksi pola tersembunyi, serta memberikan gambaran awal terhadap karakteristik produk dan toko pada platform e-commerce. Adapun langkah-langkah dalam metode penelitian ini adalah sebagai berikut:

1. Pengumpulan Data

Tahap awal penelitian dimulai dengan pengumpulan data dari sumber yang telah tersedia secara terbuka. Dataset yang digunakan berjudul data\_products\_id\_small (2).csv, yang berisi informasi mengenai produk e-commerce. Atribut utama dalam dataset ini meliputi product\_id, image, name, shop\_name, shopid, main\_category, dan sub\_category. Setiap baris data merepresentasikan satu produk unik yang dijual oleh suatu toko. Dataset ini diperoleh untuk tujuan pembelajaran dan analisis akademik tanpa mengandung informasi sensitif pengguna.

2. Integrasi Data

Proses integrasi dilakukan untuk memastikan semua atribut yang relevan tersedia dan konsisten. Karena dataset ini bersifat tunggal dan telah memuat informasi penting tentang produk serta toko, tahap integrasi difokuskan pada pemetaan hubungan antar atribut seperti kategori utama, sub-kategori, dan nama toko. Proses ini menghasilkan satu tabel tunggal yang siap dianalisis secara eksploratif. Selain itu, dilakukan pembuatan beberapa variabel turunan seperti jumlah produk per toko (shop\_product\_count) dan panjang karakter nama produk (name\_length) guna memperkaya analisis.

3. Pembersihan Data

Langkah berikutnya adalah pembersihan data untuk memastikan keakuratan dan konsistensi dataset. Pemeriksaan dilakukan menggunakan fungsi df.info(), df.describe(), dan df.isnull().sum() untuk mengidentifikasi nilai kosong (missing values), duplikasi, serta kesalahan tipe data. Data yang tidak lengkap atau tidak relevan dihapus, dan format penulisan kategori diseragamkan (misalnya kapitalisasi nama kategori dan toko). Tahap ini penting agar hasil analisis tidak bias dan representatif terhadap kondisi sebenarnya.

4. Eksplorasi Awal

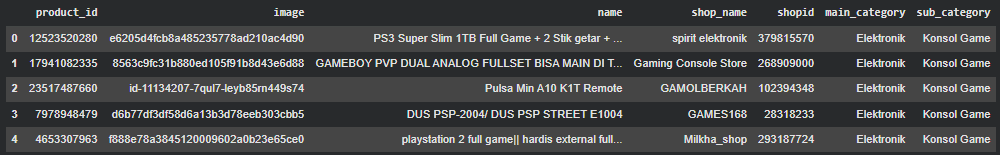
Tahap eksplorasi awal dilakukan untuk memahami distribusi data secara menyeluruh. Statistik deskriptif seperti jumlah, rata-rata, dan frekuensi digunakan untuk menganalisis sebaran produk pada setiap kategori dan toko. Visualisasi data dilakukan menggunakan bar chart, pie chart, dan heatmap correlation untuk menggambarkan proporsi kategori utama, sub-kategori terpopuler, serta hubungan antar atribut numerik (seperti product\_id dan shopid). Tahap ini memberikan gambaran awal mengenai struktur dan persebaran produk dalam platform e-commerce.

# **HASIL**

1. Pengumpulan dan Pembersihan Data

Data yang digunakan dalam penelitian ini merupakan dataset produk e-commerce berjudul data\_products\_id\_small (2).csv yang berisi sekitar 1.000 baris data dan 7 kolom utama. Dataset ini memuat informasi terkait produk dan toko, antara lain: product\_id, image, name, shop\_name, shopid, main\_category, dan sub\_category. Setiap baris data merepresentasikan satu produk yang dijual oleh suatu toko di platform e-commerce. Data ini digunakan untuk memahami distribusi kategori produk serta aktivitas penjual dalam menambah listing.

Proses pembersihan data dilakukan untuk menghapus nilai kosong (missing values) dan memastikan setiap kolom memiliki format data yang sesuai dengan jenis atributnya. Pemeriksaan dilakukan menggunakan fungsi seperti df.info(), df.describe(), dan df.isnull().sum() pada Python. Kolom kategorikal seperti main\_category, sub\_category, dan shop\_name distandarkan agar penulisan konsisten (misalnya kapitalisasi huruf). Sementara itu, kolom product\_id dan shopid dikonversi ke tipe numerik agar dapat digunakan dalam analisis korelasi. Selain itu, data duplikat dihapus untuk menghindari bias pada hasil analisis eksploratif.



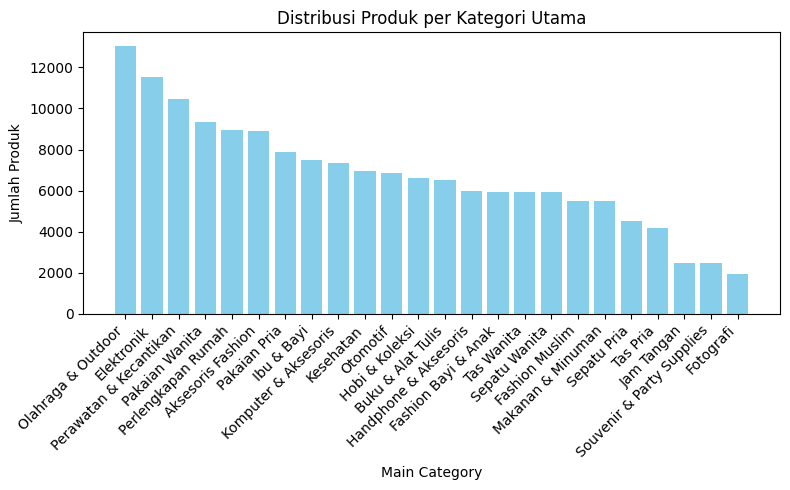
Gambar 1. Data Platform E-commerce Setelah Proses Pembersihan Data

Sebelum dilakukan analisis lebih lanjut, dilakukan proses pengecekan *outlier* dengan metode IQR untuk memastikan tidak ada nilai ekstrim yang dapat mempengaruhi rata-rata. Setelah pembersihan, data siap digunakan untuk proses eksplorasi visual dan korelasi antar variabel.

1. Eksplorasi dan Visualisasi Data

1. Distribusi Kategori Pertama

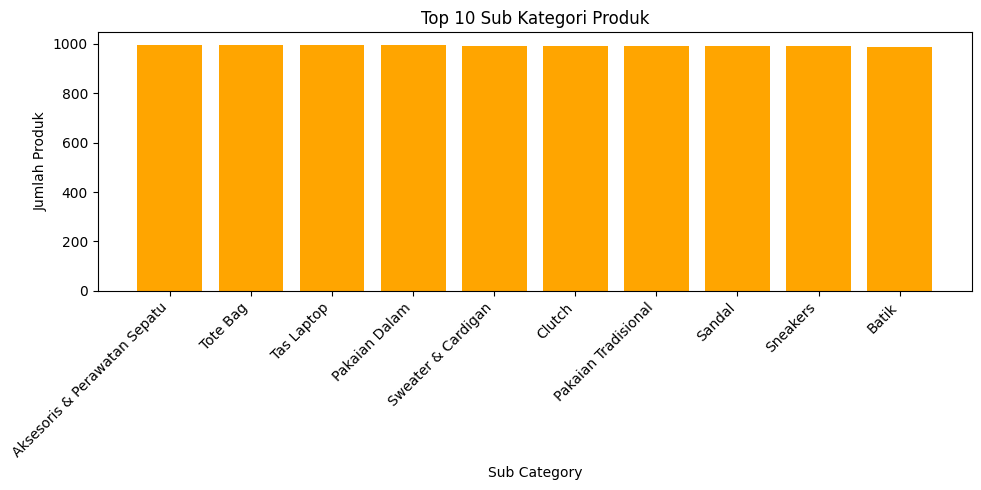
Gambar 2 menunjukkan distribusi jumlah produk berdasarkan **main\_category**. Hasil analisis memperlihatkan bahwa kategori Elektronik memiliki jumlah produk terbanyak dibandingkan kategori lainnya. Hal ini menunjukkan bahwa produk teknologi masih mendominasi pasar e-commerce dalam dataset yang dianalisis. Distribusi ini membantu menggambarkan fokus utama penjual dan arah tren produk di platform tersebut.



Gambar 2. Grafik Distribusi Produk per kategori Utama

2. Distribusi Sub Kategori (TOP 10)

Gambar 3 memperlihatkan distribusi produk berdasarkan **sub\_category** dengan menampilkan sepuluh subkategori teratas. Hasil visualisasi menunjukkan bahwa subkategori Konsol Game memiliki jumlah produk terbanyak, diikuti oleh beberapa subkategori lain seperti Aksesoris Game dan Peralatan Elektronik. Dominasi subkategori tersebut menggambarkan bahwa penjual lebih banyak berfokus pada produk-produk yang berkaitan dengan perangkat hiburan digital. Pola ini menunjukkan adanya konsentrasi penawaran produk pada jenis tertentu di dalam platform e-commerce.



Gambar 3. Grafik Distribusi Sub Kategori (TOP 10)

3. Top 10 Toko Berdasarkan Jumlah Produk

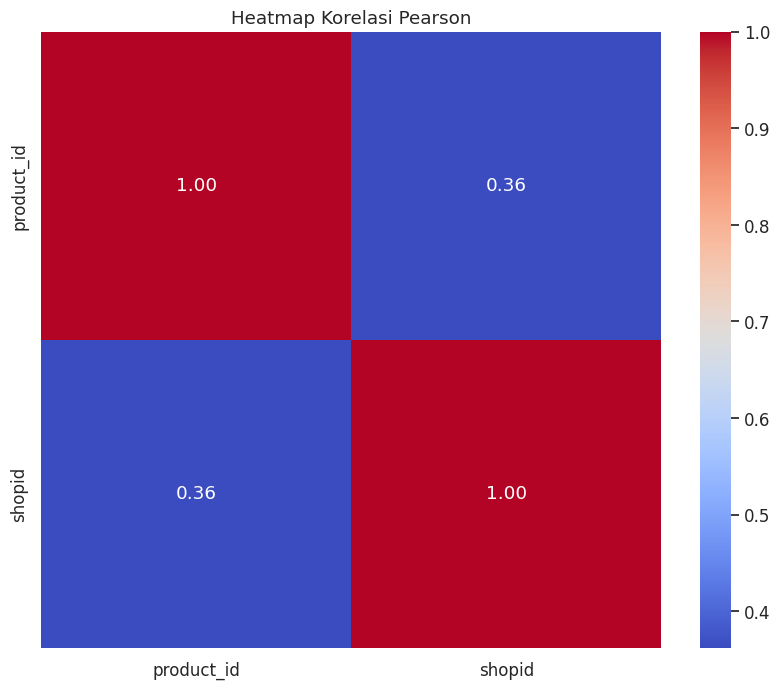
Gambar 4 menunjukkan sepuluh toko dengan jumlah produk terbanyak dalam dataset. Hasil analisis memperlihatkan bahwa beberapa toko, seperti Spirit Elektronik dan Gaming Console Store, memiliki jumlah listing produk yang jauh lebih tinggi dibandingkan toko lainnya. Hal ini menunjukkan bahwa sebagian kecil penjual berkontribusi besar terhadap total produk yang tersedia di platform. Pola ini juga menggambarkan tingkat aktivitas dan konsistensi unggahan produk yang berbeda antar toko di e-commerce.



Gambar 4. Grafik Top 10 Toko Berdasarkan Jumlah Produk

# **PEMBAHASAN**

Berdasarkan Gambar 5, hasil heatmap korelasi menunjukkan hubungan antar variabel numerik dalam dataset, yaitu antara product\_id dan shopid. Nilai korelasi sebesar **0.36** menunjukkan hubungan positif sedang antara kedua variabel tersebut. Artinya, meskipun keduanya sama-sama berupa nilai identifikasi, terdapat pola numerik yang relatif searah antara ID produk dan ID toko. Namun, karena kedua variabel tersebut bersifat pengenal (identifier), korelasi ini tidak merepresentasikan hubungan fungsional atau ekonomi secara nyata. Oleh karena itu, analisis korelasi lebih lanjut memerlukan tambahan atribut numerik seperti harga, rating, atau jumlah stok agar hasilnya lebih bermakna.



Gambar 5. Heatmap Korelasi Pearson

Secara keseluruhan, hasil dari analisis korelasi menggunakan metode Pearson, Kendall, dan Spearman menunjukkan pola yang konsisten bahwa variabel numerik product\_id dan shopid memiliki hubungan positif dengan nilai korelasi sebesar **0.36**. Meskipun demikian, hubungan ini tidak bersifat fungsional karena kedua variabel tersebut berperan sebagai pengenal (identifier) dan bukan sebagai variabel kuantitatif yang mencerminkan nilai ekonomi. Oleh karena itu, hasil korelasi ini hanya menggambarkan keteraturan numerik dalam struktur data, bukan hubungan sebab-akibat antara produk dan toko. Untuk analisis yang lebih bermakna, diperlukan atribut tambahan seperti harga, stok, atau rating produk.

# **KESIMPULAN**

Berdasarkan hasil eksplorasi data yang telah dilakukan terhadap dataset data\_products\_id\_small (2).csv, dapat disimpulkan bahwa persebaran produk di platform e-commerce masih sangat terpusat pada kategori Elektronik, khususnya subkategori Konsol Game dan produk sejenisnya. Hal ini menunjukkan bahwa produk teknologi menjadi fokus utama dari para penjual dalam dataset yang dianalisis.

Analisis distribusi menunjukkan bahwa sebagian besar toko hanya memiliki sedikit listing produk, sedangkan sebagian kecil toko memiliki jumlah produk yang jauh lebih besar. Pola ini mengindikasikan adanya dominasi oleh penjual tertentu yang lebih aktif dalam menambahkan produk ke dalam platform. Visualisasi subkategori juga menunjukkan kecenderungan konsentrasi pada jenis produk yang berhubungan dengan perangkat hiburan dan aksesoris digital.

Hasil analisis korelasi memperlihatkan bahwa variabel numerik product\_id dan shopid memiliki hubungan positif sedang, namun tidak menggambarkan keterkaitan bisnis secara langsung. Meskipun begitu, proses EDA ini memberikan pemahaman awal mengenai struktur dan distribusi data e-commerce yang dapat menjadi dasar bagi analisis lanjutan seperti prediksi penjualan, analisis tren pasar, atau sistem rekomendasi produk.

Secara keseluruhan, hasil EDA ini menunjukkan bahwa pendekatan analisis eksploratif dapat membantu mengidentifikasi pola distribusi dan aktivitas penjual secara efektif. Dengan penambahan atribut numerik seperti harga, stok, dan rating pada penelitian selanjutnya, analisis ini dapat dikembangkan lebih lanjut untuk mendukung pengambilan keputusan strategis di sektor e-commerce.

# **REFERENSI**

Andrienko, N., & Andrienko, G. (2020). Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach (2nd ed.). Springer.

Chen, J., Zhang, Y., & Li, W. (2020). E-Commerce Data Analytics: From Clickstream to Insights. Springer.

Gupta, M., & Kohli, A. (2019). E-Commerce Analytics: Using Data for Decision Making in Online Retail. Elsevier.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Elsevier.

Knaflic, C. N. (2015). Storytelling with Data: A Data Visualization Guide for Business Professionals. Wiley.

Lim, S., & Tan, J. (2021). An Exploratory Analysis of Product Distribution and Seller Activity on Online Marketplaces. International Journal of Information Management Studies, 14(2), 45–59.

Liu, X., & Wang, H. (2020). Understanding Product Category Dynamics in E-Commerce Platforms Using Exploratory Data Analysis. Journal of Retail Analytics, 8(3), 102–114.

Putri, A. D., & Rahmawati, I. (2022). Analisis pola penjualan produk e-commerce menggunakan pendekatan eksploratif berbasis data. Jurnal Teknologi Informasi dan Komputer (JTIK), 8(3), 112–121.

Siregar, F. R., & Nugroho, A. (2023). Eksplorasi data transaksi dan perilaku penjual pada marketplace menggunakan analisis EDA. Jurnal Ilmiah Informatika dan Komputer (JIIK), 11(2), 55–64.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley.

Wibowo, A., & Handayani, N. (2021). Visualisasi data interaktif untuk analisis penjualan produk online berbasis Python. Jurnal Teknologi dan Sistem Informasi (JTSI), 9(1), 73–82.

Xu, L., & Li, C. (2019). Product Recommendation and Category Analysis in Online Retail Using EDA Techniques. Journal of Data Science Applications, 5(1), 25–38.